

Linear Tail-Biting Trellises: Characteristic Generators and the BCJR-Construction

Heide Gluesing-Luerssen* and Elizabeth Weaver

University of Kentucky
 Department of Mathematics
 715 Patterson Office Tower
 Lexington, KY 40506-0027, USA;
 {heidegl,eweaver}@ms.uky.edu

March 23, 2010

Abstract: We investigate the constructions of tail-biting trellises for linear block codes introduced by Koetter/Vardy [9] and Nori/Shankar [15]. For a given code we will define the sets of characteristic generators more generally than in [9] and we will investigate how the choice of characteristic generators affects the set of resulting product trellises, called KV-trellises. Furthermore, we will show that each KV-trellis is a BCJR-trellis, defined in a slightly stronger sense than in [15], and that the latter are always non-mergeable. Finally, we will address a duality conjecture of Koetter/Vardy by making use of a dualization technique of BCJR-trellises and prove the conjecture for minimal trellises.

Keywords: linear block codes, tail-biting trellises, characteristic generators, tail-biting BCJR-trellises, minimal trellises

MSC (2000): 94B05, 94B12, 68R10, 93B20

1 Introduction

During the last decade conventional and tail-biting trellis representations of linear block codes have gained a great deal of attention because both types of trellises do not only reveal code structure but also may give rise to efficient decoding algorithms of Viterbi type.

As a consequence, the quest is on for constructing trellises with low complexity with respect to various measures. For conventional trellises it is by now well-known that for a given linear block code (and with a fixed coordinate ordering) there exists a unique minimal trellis and minimality coincides with bipartiteness as well as with non-mergeability; see [14, 12, 4]. This unique minimal trellis simultaneously minimizes the vertex count at each time as well as any other conceivable complexity measure. Several methods exist to construct this minimal trellis, the BCJR-construction [1, 12], the Forney-construction based in past-future induced realizations [4] and others. For an excellent overview on conventional trellises we refer to Vardy's survey [18].

For tail-biting trellises the situation is considerably more difficult. Due to an additional degree of freedom (the cardinality of the vertex space at time 0) minimal tail-biting trellises

*The author was partially supported by National Science Foundation grant #DMS-0908379

are not unique and this applies to minimality with respect to various measures, e. g., the state complexity profile, the maximum state dimension, the sum of the state dimensions as well as to all these notions where we replace the state spaces by the edge spaces, see, for instance, [2, 9]. Moreover, minimality, non-mergeability, and biproperness are not equivalent for tail-biting trellises. The fact, however, that the complexity of a tail-biting trellis may be much lower than that of the minimal conventional trellis raised a great deal of interest in these trellises and, in particular, in the question how to efficiently construct and to analyze tail-biting trellises; see also [17, 11] in addition to the aforementioned references.

A major breakthrough has been made by Koetter/Vardy in the excellent paper [9]. Therein, Koetter/Vardy presented a construction from which all minimal tail-biting trellises can be derived. Their construction accounts for all cyclic shifts of the code and results in a list of characteristic generators (codewords associated with spans), which then comprise what they call the characteristic matrix. It can be shown that each minimal trellis (with respect to any of the complexity measures mentioned above) is a KV-trellis, that is, arises as the product of the elementary trellises associated with characteristic generators. However, some care needs to be exercised at this point because not all minimal trellises of the given code can be obtained from the same list of characteristic generators. We will present an example later on in 3.15. Furthermore, as observed in [9], not all KV-trellises are minimal trellises. But, as we will see later on, they are always non-mergeable.

A different construction of tail-biting trellises has been introduced by Nori/Shankar in [15]. Their construction generalizes the well-known BCJR-construction of conventional trellises. The additional degree of freedom, the choice of the vertex space at time 0, is made explicit by a displacement matrix which can be regarded as capturing the circular past of the various trajectories (cycles) through the trellis. In the conventional case this past is simply zero. While the displacement matrix can be chosen arbitrarily, the best trellises arise when the matrix is based on the choice of spans for the given basis of the code. Later on we will restrict ourselves mainly to those displacement matrices. Nori/Shankar [15] showed that a non-mergeable product trellis is isomorphic to the BCJR-trellis based on the same span list. A particularly nice feature of the tail-biting BCJR-trellises is that they naturally give rise to a dualization technique, resulting in a trellis for the dual code with the same state complexity profile.

In this paper we will discuss the relation between the Koetter/Vardy and Nori/Shankar approach and present further properties. We will show that BCJR-trellises (in a slightly stronger sense than in [15]) are non-mergeable and that all KV-trellises are BCJR-trellises, hence non-mergeable. We will further show that any one-to-one product trellis can be merged to a BCJR-trellis and merging is accomplished by taking suitable quotients of the vertex spaces. Finally, we will discuss a duality conjecture stated in [9]. It deals with a particular way of dualizing a KV-trellis. This leads to a trellis with the same state complexity profile, but for which it is not a priori clear whether it represents the dual code. We will show that, in general, this depends on the choice of the characteristic matrix and we will prove that for minimal trellises there always exists a suitable choice of the dual characteristic matrix such that the conjecture holds true. In that case the dual trellis obtained via Koetter/Vardy's characteristic matrix turns out to be the same as the BCJR-dual. The main tool for the proof is a dualization technique going back to Mittelholzer [13] and Forney [5] which amounts to dualizing the edge spaces with respect to a particular bilinear form. As the proof will show, for minimal trellises the resulting edge space dual coincides with the BCJR-dual. We will conclude the paper with a list of open problems.

2 Preliminaries

In this section we introduce the basic notions for tail-biting trellises. Throughout the paper let \mathbb{F} be a finite field.

A *tail-biting trellis* $T = (V, E)$ over \mathbb{F} of depth n is a directed edge-labeled graph with the property that the vertex set V partitions into n disjoint sets $V = V_0 \cup V_1 \cup \dots \cup V_{n-1}$ such that every edge in T that starts in V_i ends in $V_{i+1 \bmod n}$ and where the edges are labeled with field elements from \mathbb{F} . Thus, the edge set E decomposes into $E = \bigcup_{i=0}^{n-1} E_i$, where E_i is the set of edges starting in V_i and ending in $V_{i+1 \bmod n}$. A *cycle* in T is a closed path of length n in the trellis. We will always assume that the cycles start and end (at the same point) in V_0 . If $|V_0| = 1$, the trellis is called *conventional* (we will not consider conventional trellises where $|V_0| > 1$). We call the trellis *reduced* if every vertex and every edge appears in at least one cycle. The trellis is called *biprime* if any two edges starting at the same vertex or ending at the same vertex are labeled distinctly. The interval $\mathcal{I} := \{0, \dots, n-1\}$ is the *time axis* of the trellis.

In the sequel we will always identify the edges with their triples consisting of starting point, label, and endpoint point. Thus, the edge sets are given by

$$E_i = \{(v, a, \hat{v}) \mid \text{there exist an edge } v \xrightarrow{a} \hat{v} \text{ where } v \in V_i, \hat{v} \in V_{i+1}, a \in \mathbb{F}\}, i \in \mathcal{I}.$$

Hence $E_i \subseteq V_i \times \mathbb{F} \times V_{i+1}$ and (V, E) completely specifies the trellis. The edge sets have also been called trellis section [2, 6] or local constraints [5].

Notice that we have to compute modulo n on the time axis \mathcal{I} . If the trellis is conventional, however, there is no need in doing so because in that case all paths are cycles and the endpoints of edges in E_{n-1} do not carry any information.

Each tail-biting trellis gives rise to its edge-label code consisting of the edge-label sequences of all cycles in T . We will need to make this and other related notions precise.

Definition 2.1 Let $T = (V, E)$ be a tail-biting trellis over \mathbb{F} of depth n and put $\mathcal{W} := V_0 \times \mathbb{F} \times \dots \times V_{n-1} \times \mathbb{F}$.

(a) The *label code* of T is defined as

$$\begin{aligned} \mathcal{S}(T) := \{(v_0, c_0, v_1, \dots, v_{n-1}, c_{n-1}) \in \mathcal{W} \mid \\ \text{there exist a cycle } v_0 \xrightarrow{c_0} v_1 \xrightarrow{c_1} \dots v_{n-1} \xrightarrow{c_{n-1}} v_0 \text{ in } T\}. \end{aligned}$$

The trellis T is called *linear* if T is reduced and the vertex sets $V_i, i \in \mathcal{I}$, are vector spaces over \mathbb{F} and $\mathcal{S}(T)$ is a subspace of \mathcal{W} . In this case the edge spaces $E_i, i \in \mathcal{I}$, are subspaces of $V_i \times \mathbb{F} \times V_{i+1}$ because they are projections of $\mathcal{S}(T)$.

(b) The *edge-label code* of T is defined as

$$\mathcal{C}(T) := \{(c_0, \dots, c_{n-1}) \in \mathbb{F}^n \mid \text{there exist a cycle } v_0 \xrightarrow{c_0} v_1 \xrightarrow{c_1} \dots v_{n-1} \xrightarrow{c_{n-1}} v_0 \text{ in } T\}.$$

We say that T represents the code $\mathcal{C} \subseteq \mathbb{F}^n$ if $\mathcal{C}(T) = \mathcal{C}$. Note that if T is linear then $\mathcal{C}(T)$ is a linear block code. The trellis T is called *one-to-one* if there exists a bijection between $\mathcal{C}(T)$ and the cycles in T . That is, every word in $\mathcal{C}(T)$ appears as the edge label sequence of exactly one cycle.

(c) If T is linear the *state complexity profile* (SCP) is defined as (s_0, \dots, s_{n-1}) , where $s_i = \dim V_i$. The *edge complexity profile* (ECP) is defined as (e_0, \dots, e_{n-1}) , where $e_i = \dim E_i$.

- (d) Two trellises $T = (V, E)$ and $T' = (V', E')$ are called *isomorphic* if there exists a bijection $\phi : V \rightarrow V'$ such that $\phi(V_i) = V'_i$ for all $i \in \mathcal{I}$ and such that $(v, a, w) \in E_i$ if and only if $(\phi(v), a, \phi(w)) \in E'_i$. If T and T' are linear, then we also require $\phi|_{V_i} : V_i \rightarrow V'_i$ be isomorphisms. Isomorphic trellises represent the same code.
- (e) Two trellises $T = (V, E)$ and $T' = (V', E')$ are called *structurally isomorphic* if there exists a bijection $\phi : V \rightarrow V'$ such that $\phi(V_i) = V'_i$ for all $i \in \mathcal{I}$ and such that for all $(v, w) \in V_i \times V_{i+1}$ the number of edges from $v \in V_i$ to $w \in V_{i+1}$ equals the number of edges from $\phi(v) \in V'_i$ to $\phi(w) \in V'_{i+1}$ (thus we disregard the edge labels). Again, if T and T' are linear we also require $\phi|_{V_i} : V_i \rightarrow V'_i$ be isomorphisms.

Throughout the paper we will only deal with linear block codes and linear (hence reduced) trellis representations. Observe that, in general, structurally isomorphic trellises do not represent the same code. However, we will only be interested in this notion for trellises of a fixed code.

We wish to briefly touch upon all these notions from the viewpoint of behavioral systems theory. A trellis T can be regarded as a dynamical system with latent variables in the sense of [16, Def. 1.3.4], where the latent variables represent the vertex sequence and satisfies the axiom of state at each time i , see [19, Def. 1.3]. In particular, the vertex space V_i is the state space at time i . In this context the edge spaces, consisting of all state-output-next state triples, are called the evolution law [19, Def. 1.4]. Completeness as defined in [19, p. 184], a crucial notion in behavioral systems theory relating local and global behavior does not play a role in our setting since, due to the finite time axis, each system is complete. Notice that the system is, in general, highly time-varying since even the state space dimensions depend on time. The label code $\mathcal{S}(T)$ is — up to the ordering of labels and vertices — the full behavior¹ while the edge-label code $\mathcal{C}(T)$ is the manifest behavior, see again [16, Def. 1.3.4]. If we consider the code \mathcal{C} as the given system, which, as usual in behavioral systems theory, is identified with its set of trajectories (the codewords), then T is simply a state-space realization of \mathcal{C} if $\mathcal{C}(T) = \mathcal{C}$. Furthermore, isomorphic trellises are equivalent systems in the sense of [19, p. 205]. They merely differ by the labeling of the latent variable (the vertices). The system theoretic notion of minimal realizations now becomes minimal trellises, that is, minimizing the amount of data needed for a realization. While for conventional trellises this is well-understood and amounts to the unique minimal trellis (up to isomorphism) with a variety of universal properties, the situation is much more complicated for tail-biting trellises. This leads to various notions of minimality for trellises of which only the following one will be relevant for us. For other notions see for instance [14, 12] or [9, p. 2085]. In this context we will also introduce the notion of mergeability of trellises (In system theoretic language merging as defined below is also referred to as lumping states, see [19, p. 207]).

Definition 2.2 Let $\mathcal{C} \subseteq \mathbb{F}^n$ be a linear block code and $T = (V, E)$ be a linear trellis of \mathcal{C} .

- (a) T is called *minimal* if there exists no linear trellis $T' = (V', E')$ of \mathcal{C} such that $|V'_i| \leq |V_i|$ for all $i \in \mathcal{I}$ and $|V'_j| < |V_j|$ for some $j \in \mathcal{I}$.
- (b) T is called *non-mergeable* if for any $i \in \mathcal{I}$ merging any two vertices $v_1, v_2 \in V_i$ results in a trellis \hat{T} that does not represent \mathcal{C} . Here merging means replacing the two vertices v_1, v_2 in the trellis by a single vertex $\hat{v} \in V_i$ and replacing each edge of the form $(v_j, a, w) \in E_i$, resp. $(w, a, v_j) \in E_{i-1}$, where $j = 1, 2$, by the edge $(\hat{v}, a, w) \in E_i$, resp. $(w, a, \hat{v}) \in E_{i-1}$.

Due to our restriction to linear trellises it seems necessary to also address the issue whether

¹Strictly speaking, the full behavior is the space of all paths in T and not just the cycles.

merging leads to a potentially non-linear trellis. It turns out however, that mergeability is compatible with the linear structure and hence merging can always be accomplished such that the resulting trellis is linear again.

It is not a priori clear whether minimal (linear) trellises exist for a given linear code. For conventional trellises this is indeed the case as is well-known since a long time and leads to minimizing many other measures as well; see [18, Thm. 4.26]. Moreover, all minimal conventional trellises are isomorphic and even minimal in the larger class of *all* trellises, not necessarily linear. Furthermore, for conventional trellises minimality coincides with non-mergeability. A nice way of constructing the unique minimal conventional trellis of a given code, using realization theory for dynamical systems, has been given by Forney in [4]. For tail-biting trellises minimality has been studied in much detail by Calderbank/Forney/Vardy in [2] and Koetter/Vardy in [9]. In particular, it has been shown that linear minimal tail-biting trellises exist (the minimal conventional trellis is one of them) and there exist non-isomorphic ones. This is actually easy to see by constructing the minimal conventional trellis for each cyclic shift of the given code and then shifting back that trellis; see [9, p. 2084]. This results in many (non-isomorphic) minimal trellises and in each one at least one vertex space V_i is trivial. However, in general there exist other minimal trellises (that is, for which none of the vertex spaces is trivial); see for instance [9, Exa. 6.1] for the $(8, 4, 4)$ Hamming code, which also appeared in [2, p. 1449]. Moreover, different from the situation for conventional trellises, minimality is stronger than non-mergeability for tail-biting trellises; see for instance [9, Exa. 3.3]. The main result of [9] is a construction from which all minimal tail-biting trellises of a given code can be derived. Our paper is mainly devoted to investigating this construction, which we will therefore introduce in the next section.

Finally, let us fix the following notation. For a matrix $M \in \mathbb{F}^{m \times n}$ we denote its row span as $\text{im } M$, thus $\text{im } M := \{\alpha M \mid \alpha \in \mathbb{F}^m\}$. If not otherwise stated, $\mathcal{C} \subseteq \mathbb{F}^n$ denotes a k -dimensional code of length n over the finite field \mathbb{F} with the following generator and parity check matrices:

$$\mathcal{C} = \text{im } G, \text{ where } G = (g_{lj}) = \begin{pmatrix} g_1 \\ \vdots \\ g_k \end{pmatrix} = (G_0 \quad \dots \quad G_{n-1}) \in \mathbb{F}^{k \times n}, \quad (2.1)$$

and

$$\mathcal{C} = \ker H^T := \{c \in \mathbb{F}^n \mid cH^T = 0\}, \text{ where } H^T = \begin{pmatrix} H_0 \\ \vdots \\ H_{n-1} \end{pmatrix} \in \mathbb{F}^{n \times (n-k)}. \quad (2.2)$$

Hence $G_i \in \mathbb{F}^k$ and $H_i \in \mathbb{F}^{n-k}$ are the columns of G and rows of H^T , respectively, while $g_1, \dots, g_k \in \mathbb{F}^n$ are the rows of G . For simplicity we will also assume that the support of \mathcal{C} is $\mathcal{I} = \{0, \dots, n-1\}$, that is, no column of G is zero. Finally, it will be convenient to have a notion for the indicator function of a subset $\mathcal{A} \subseteq \mathcal{I}$. Thus, we define $I^{\mathcal{A}} \in \mathbb{F}^n$, where $I_j^{\mathcal{A}} = 1$ if $j \in \mathcal{A}$ and $I_j^{\mathcal{A}} = 0$ else.

3 Product Trellises and KV-Trellises

We begin with recalling the construction of tail-biting trellises based on products of elementary trellises. The underlying notions and constructions are well-known, see [10] for the conventional case and [2, 9] for the tail-biting case. Thereafter, we will introduce the notion of a characteristic pair for a given code, from which in essence all minimal tail-biting trellises

can be derived. This crucial and very fruitful concept goes back to Koetter/Vardy [9], who in turn based it on previous work by Kschischang/Sorokine [10]. The relation between minimal trellises and characteristic matrices will be discussed in detail.

Due to the cyclic structure of the time axis \mathcal{I} for tail-biting trellises, the following interval notation has proven to be very convenient. For $a, b \in \mathcal{I}$ we define

$$\begin{aligned} [a, b] &= \left\{ \begin{array}{ll} \{a, a+1, \dots, b\} & \text{if } a \leq b, \\ \{a, a+1, \dots, n-1, 0, 1, \dots, b\} & \text{if } a > b, \end{array} \right. \\ (a, b] &= [a, b] \setminus \{a\} \end{aligned} \quad \text{for all } a, b. \quad (3.1)$$

We call the intervals $(a, b]$ and $[a, b]$ *linear* if $a \leq b$ and *circular* else. Notice that $(a, a] = \emptyset$. It is easy to see that

$$\mathcal{I} \setminus (a, b] = (b, a] \text{ for all } a \neq b, \quad (3.2)$$

hence the complement of a nonempty linear interval is circular and vice versa. Observe also that $0 \notin (a, b] \iff (a, b]$ is linear.

Definition 3.1 For a nonzero vector $c = (c_0, \dots, c_{n-1}) \in \mathbb{F}^n$ we call the interval $(a, b]$ a *span* of c if $c_a \neq 0 \neq c_b$ and $c_j = 0$ for $j \notin [a, b]$. In this case we will also call $[a, b]$ the *closed span* of c .

For instance, the vector $v = (0, 1, 3, 0, 2, 2) \in \mathbb{F}_5^6$ has the spans $(1, 5]$, $(2, 1]$, $(4, 2]$, and $(5, 4]$. Notice that every nonzero vector has a unique linear span. It turns out that not including the starting point a in the span of x is very convenient for later notation; see, for instance, the next definition.

Definition 3.2 Let $c = (c_0, \dots, c_{n-1}) \in \mathbb{F}^n$ and $(a, b]$ be a span of c . The *elementary trellis for the pair $(c, (a, b])$* is defined as $T_{c, (a, b]} := (V, E)$, where the vertex sets and edge sets $V = \bigcup_{i=0}^{n-1} V_i$ and $E = \bigcup_{i=0}^{n-1} E_i$ are as follows:

if $a = b$ then,

$$V_i := \{0\} \text{ for all } i \in \mathcal{I} \text{ and } E_i := \left\{ \begin{array}{ll} \{(0, 0, 0)\}, & \text{if } i \neq a \\ \{(0, \alpha c_i, 0) \mid \alpha \in \mathbb{F}\}, & \text{if } i = a; \end{array} \right.$$

if $a \neq b$, then

$$V_i := \begin{cases} 0, & \text{if } i \notin (a, b] \\ \mathbb{F}, & \text{if } i \in (a, b] \end{cases} \quad \text{and} \quad E_i := \left\{ \begin{array}{ll} \{(0, 0, 0)\}, & \text{if } i \notin [a, b] \\ \{(0, \alpha c_i, \alpha) \mid \alpha \in \mathbb{F}\}, & \text{if } i = a \\ \{(\alpha, \alpha c_i, 0) \mid \alpha \in \mathbb{F}\}, & \text{if } i = b \\ \{(\alpha, \alpha c_i, \alpha) \mid \alpha \in \mathbb{F}\}, & \text{if } i \in (a, b - 1] \end{array} \right.$$

Using the indicator function $I^{(a, b]}$, the above vertex and edge spaces can be written as $V_i = \text{im}(I_i^{(a, b)}) \subseteq \mathbb{F}$ and $E_i = \text{im}(I_i^{(a, b)}, c_i, I_{i+1}^{(a, b)}) \subseteq V_i \times \mathbb{F} \times V_{i+1}$.

Note that $T_{c, (a, b]}$ is a one-to-one linear trellis and $\mathcal{C}(T) = \text{im } c = \{\alpha c \mid \alpha \in \mathbb{F}\}$, the one-dimensional code generated by c .

With the aid of the product construction as introduced in [10, 9] we can derive trellises for higher-dimensional codes. Recall that the product $\hat{T} := T \times T'$ of two trellises $T = (V, E)$ and $T' = (V', E')$ of depth n over \mathbb{F} is defined as the trellis $\hat{T} = (\hat{V}, \hat{E})$, where $\hat{V}_i := V_i \times V'_i$

and $\hat{E}_i = \{(v, v'), a + a', (w, w') \mid (v, a, w) \in E_i, (v', a', w') \in E'_i\}$ for all $i \in \mathcal{I}$. The product trellis \hat{T} satisfies $\mathcal{C}(\hat{T}) = \mathcal{C}(T) + \mathcal{C}(T')$.

The most important application of this construction is the product of elementary trellises leading to tail-biting trellis representations of linear block codes. Recall the data for the given code $\mathcal{C} \subseteq \mathbb{F}^n$ as given in (2.1).

Theorem 3.3 ([10, 8, 9]) *Let*

$$\mathcal{S} := [(a_l, b_l], l = 1, \dots, k] \quad (3.3)$$

be a span list for G , that is, $(a_l, b_l]$ is a span (linear or circular) for the row g_l , $l = 1, \dots, k$. Then product trellis $T_{G, \mathcal{S}} := T_{g_1, (a_1, b_1]} \times \dots \times T_{g_k, (a_k, b_k]}$ represents the code \mathcal{C} , that is, $\mathcal{C}(T_{G, \mathcal{S}}) = \mathcal{C}$. Moreover, $T_{G, \mathcal{S}}$ is linear (in particular reduced) and one-to-one. The vertex spaces of $T_{G, \mathcal{S}}$ are given by

$$V_i = \text{im } M_i, \text{ where } M_i = \begin{pmatrix} \mu_i^1 & & \\ & \ddots & \\ & & \mu_i^k \end{pmatrix} \in \mathbb{F}^{k \times k} \text{ and } \mu_i^l = \begin{cases} 1, & \text{if } i \in (a_l, b_l] \\ 0, & \text{if } i \notin (a_l, b_l] \end{cases}, \quad (3.4)$$

while the edge spaces are given by $E_i = \text{im } (M_i, G_i, M_{i+1})$ for $i \in \mathcal{I}$.

Observe that $\mu^l = I^{(a_l, b_l]}$ is the indicator function of $(a_l, b_l]$. It is straightforward to see that, more generally, a product trellis $T := T_{g_1, (a_1, b_1]} \times \dots \times T_{g_k, (a_k, b_k]}$ is one-to-one if and only if g_1, \dots, g_k are linearly independent.²

In the sequel we will only consider one-to-one product trellises (with the exception of Remark 4.15). In addition to this restriction our product trellises are confined even further when compared to the class considered by Koetter/Vardy [9]. Due to Definition 3.1 we only consider “shortest spans” for the generators, that is, the vector is nonzero at the endpoints of the span (in other words, the closed span is a smallest linear or circular interval containing the support). This restriction is justified by the fact that an elementary trellis based on a non-shortest span is mergeable and thus so is the product trellis, see [9, Lemma 4.3]. Of course, even with shortest spans a product trellis may be mergeable; we will see examples later on.

The following shift property of product trellises is a direct consequence. Recall that we compute with indices modulo n . This, of course, also applies to the span lists.

Remark 3.4 Let $\sigma : \mathbb{F}^n \rightarrow \mathbb{F}^n$, $(c_0, \dots, c_{n-1}) \mapsto (c_1, \dots, c_{n-1}, c_0)$ be the left shift on \mathbb{F}^n and let $G^* \in \mathbb{F}^{k \times n}$ be the matrix consisting of the shifted rows $\sigma(g_l)$, $l = 1, \dots, k$. If \mathcal{S} as in (3.3) is a span list for G then $\mathcal{S}^* = [(a_l - 1, b_l - 1], l = 1, \dots, k]$ forms a span list for G^* . Furthermore, with the notation as in (3.4) the vertex spaces of the product trellis T_{G^*, \mathcal{S}^*} are given by $V_i^* = \text{im } M_{i+1}$ and the edge spaces are given by $E_i^* = \text{im } (M_{i+1}, G_{i+1}, M_{i+2})$ for $i \in \mathcal{I}$.

For the product trellises $T_{G, \mathcal{S}}$ it is not hard to give formulas for the SCP and the ECP in terms of the span list \mathcal{C} . Let $\underline{k} := \{1, \dots, k\}$.

²There is one extreme case to exclude from this statement: if some spans are of the form $(a_l, a_l]$ then T is one-to-one if and only if the set \mathcal{G} is linearly independent, where \mathcal{G} is obtained by taking for each span of the form $(a_l, a_l]$ only one generator and taking all generators with non-empty span. Since in our considerations g_1, \dots, g_k are always linearly independent this case will not be relevant for us.

Proposition 3.5 Let \mathcal{S} and $T_{G,\mathcal{S}}$ be as in Theorem 3.3. Put $\mathcal{A} := \{a_1, \dots, a_k\}$ and $\mathcal{B} = \{b_1, \dots, b_k\}$. Moreover, for each $i \in \mathcal{I}$ define

$$\mathcal{L}_i := \{l \in \underline{k} \mid i \in (a_l, b_l]\}. \quad (3.5)$$

Then the SCP of $T_{G,\mathcal{S}}$ is given by (s_0, \dots, s_{n-1}) , where $s_i = |\mathcal{L}_i|$. The ECP (e_0, \dots, e_{n-1}) satisfies the following.

- (a) If a_1, \dots, a_k are distinct, then $e_i = s_i + I_i^{\mathcal{A}}$ for $i \in \mathcal{I}$.
- (b) If b_1, \dots, b_k are distinct, then $e_i = s_{i+1} + I_i^{\mathcal{B}}$ for $i \in \mathcal{I}$.
- (c) Let a_1, \dots, a_k be distinct and b_1, \dots, b_k be distinct. Then the SCP satisfies

$$s_{i+1} = \begin{cases} s_i, & \text{if } i \in \mathcal{A} \cap \mathcal{B} \text{ or } i \notin \mathcal{A} \cup \mathcal{B}, \\ s_i + 1, & \text{if } i \in \mathcal{A} \setminus \mathcal{B}, \\ s_i - 1, & \text{if } i \in \mathcal{B} \setminus \mathcal{A}, \end{cases}$$

and the trellis $T_{G,\mathcal{S}}$ is biproper.

Observe that if all spans $(a_l, b_l]$ are linear, then the condition a_1, \dots, a_k and b_1, \dots, b_k both being distinct is equivalent to G being an MSGM in the sense of [12, Def. 6.2, Thm. 6.11] (also called ‘‘sets of atomic generators’’ in [10] or ‘‘shortest basis’’ in [3]). Moreover, in this case $T_{G,\mathcal{S}}$ is the unique minimal conventional trellis of \mathcal{C} and the formulas given in (a) and (b) coincide with those known for conventional trellises; see, for instance, [12, p. 1080]. However, it is well-known that if not all spans are linear the (tail-biting) product trellis is in general not minimal; see, for instance, Example 4.10 later on.

PROOF: The first statement about the SCP is immediate from the definition of the vertex spaces $V_i = \text{im } M_i$ in (3.4). For the ECP, recall that by definition of the edge spaces we have $e_i = \text{rk}(M_i, G_i, M_{i+1})$.

(a) We proceed as follows. For a matrix $M \in \mathbb{F}^{k \times r}$ denote by $\text{col}(M) \subseteq \mathbb{F}^k$ its column space and by $\text{col}(M, t)$ its t -th column. Fix $i \in \mathcal{I}$. Let $\hat{e}_i := \text{rk}(M_i, G_i)$. In a first step we will show that $\hat{e}_i = s_i + I_i^{\mathcal{A}}$ and in a second step we will prove that $e_i = \hat{e}_i$. For the first step, notice that $\hat{e}_i = s_i$ if $G_i \in \text{col}(M_i)$ and $s_i + 1$ else. Moreover, the definition of the matrices M_i implies

$$G_i \in \text{col}(M_i) \iff g_{li} = 0 \text{ for all } l \notin \mathcal{L}_i.$$

Since $(a_l, b_l]$ is a span for the row g_l one has $g_{li} = 0$ for all $i \notin [a_l, b_l]$ and $g_{l,a_l} \neq 0$. But then the above along with $(a_l, b_l] \cup \{a_l\} = [a_l, b_l]$ shows that $G_i \in \text{col}(M_i) \iff i \notin \mathcal{A}$. All this proves $\hat{e}_i = s_i + I_i^{\mathcal{A}}$. In order to establish $e_i = \hat{e}_i$ it remains to show that $\text{col}(M_{i+1}) \subseteq \text{col}(M_i, G_i)$. From (3.4) we know that $\text{col}(M_{i+1}, l) = 0$ if $l \notin \mathcal{L}_{i+1}$ and $\text{col}(M_{i+1}, l) = \varepsilon_l$ if $l \in \mathcal{L}_{i+1}$, where $\varepsilon_l \in \mathbb{F}^k$ denotes the k -th standard basis vector. Hence fix $l \in \mathcal{L}_{i+1}$. If $l \in \mathcal{L}_i$, then $\text{col}(M_{i+1}, l) = \varepsilon_l = \text{col}(M_i, l) \in \text{col}(M_i, G_i)$ and we are done. If $l \notin \mathcal{L}_i$, then $i \notin (a_l, b_l]$ and thus $\text{col}(M_i, l) = 0$. But then the assumption $i+1 \in (a_l, b_l]$ implies that $i = a_l$ and thus $g_{li} \neq 0$. In order to examine the other entries of the column vector G_i let $r \in \underline{k} \setminus \{l\}$. If $r \in \mathcal{L}_i$ then $\text{col}(M_i, r) = \varepsilon_r$. If $r \notin \mathcal{L}_i$ then $i \notin (a_r, b_r]$ and since $i = a_l$, which is distinct from a_r , this implies $i \notin [a_r, b_r]$, hence $g_{ri} = 0$. All this shows that $G_i - \sum_{r=1}^k g_{ri} \text{col}(M_i, r) = g_{li} \varepsilon_l = g_{li} \text{col}(M_{i+1}, l)$. As a consequence, $\text{col}(M_{i+1}, l) \in \text{col}(M_i, G_i)$ and therefore $e_i = \text{rk}(M_i, G_i, M_{i+1}) = \text{rk}(M_i, G_i) = s_i + I_i^{\mathcal{A}}$.

(b) We proceed similarly using the rank $\tilde{e}_i := \text{rk}(G_i, M_{i+1})$. This time we have

$$\begin{aligned} G_i \in \text{col}(M_{i+1}) &\iff g_{li} = 0 \text{ for all } l \text{ such that } i+1 \notin (a_l, b_l] \\ &\iff g_{li} = 0 \text{ for all } l \text{ such that } i \notin (a_l - 1, b_l - 1]. \end{aligned} \quad (3.6)$$

The latter semi-open interval is $[a_l, b_l - 1]$ if $a_l \neq b_l$ and empty otherwise. Since $g_{l,b_l} \neq 0$ (3.6) implies the equivalence $G_i \in \text{col}(M_{i+1}) \iff i \notin \mathcal{B}$. Hence $\tilde{e}_i = s_{i+1} + I_i^{\mathcal{B}}$ and it remains to show that $\text{col}(M_i) \subseteq \text{col}(G_i, M_{i+1})$. Fix $l \in \mathcal{L}_i$. Then $\text{col}(M_i, l) = \varepsilon_l$. If $l \in \mathcal{L}_{i+1}$ then $\text{col}(M_{i+1}, l) = \varepsilon_l$ and we are done. If $l \notin \mathcal{L}_{i+1}$, then $i+1 \notin (a_l, b_l]$, but $i \in (a_l, b_l]$, thus $i = b_l$. In this case $\text{col}(M_{i+1}, l) = 0$ and $g_{li} \neq 0$. For the other entries of G_i pick $r \in \underline{k} \setminus \{l\}$. If $r \in \mathcal{L}_{i+1}$ then $\text{col}(M_{i+1}, r) = \varepsilon_r$. If $r \notin \mathcal{L}_{i+1}$, then $i+1 \notin (a_r, b_r]$, hence $i \notin (a_r - 1, b_r - 1]$. But since $i = b_l$ is different from b_r this implies $i \notin [a_r, b_r]$. Hence $g_{ri} = 0$. All this shows that $G_i - \sum_{r=1}^k g_{ri} \text{col}(M_{i+1}, r) = g_{li} \varepsilon_l = g_{li} \text{col}(M_i, l)$ and therefore $\text{col}(M_i, l) \in \text{col}(G_i, M_{i+1})$. In particular, $\text{rk}(M_i, G_i, M_{i+1}) = \text{rk}(G_i, M_{i+1})$.

(c) The identities for s_i follow from equating the expressions for e_i in (a) and (b). Furthermore, in (a) and (b) we saw that $\text{rk}(M_i, G_i) = \text{rk}(G_i, M_{i+1}) = \text{rk}(M_i, G_i, M_{i+1})$. This in turn implies $\ker(M_i, G_i) = \ker(G_i, M_{i+1}) = \ker(M_i, G_i, M_{i+1})$, from which the biproperness follows. \square

At this point we wish to show a path property of product trellises that will prove useful later on when discussing mergeability. The property holds true under a certain condition on the vertex spaces, satisfied by product trellises $T_{G,S}$, but also by the BCJR-trellises studied in the next section. In system theoretic language this path property means that the trellis is point controllable (in n steps) in the sense of [19, p. 188].

Lemma 3.6 *Let $\mathcal{C} = \text{im } G$ and let the span list \mathcal{S} be as in (3.3). Let $T = (V, E)$ be a trellis for \mathcal{C} such that the vertex and edge spaces are given by $V_i = \text{im } M_i$ and $E_i = \text{im}(M_i, G_i, M_{i+1})$ for certain matrices $M_i \in \mathbb{F}^{k \times r}$. Let the sets \mathcal{L}_i be defined as in (3.5) and suppose that for all $i \in \mathcal{I}$ the l -th row of M_i is zero whenever $l \notin \mathcal{L}_i$.*

Then for every $v \in V_0 = \text{im } M_0$ there exists a path through the trellis T that starts at $v \in V_0$ and ends at $0 \in V_0$. Precisely, there exist vectors $\alpha^{(0)}, \dots, \alpha^{(n-1)} \in \mathbb{F}^k$ such that

$$\alpha^{(0)} M_0 = v, \quad \alpha^{(i)} - \alpha^{(i+1)} \in \ker M_{i+1} \text{ for } i = 0, \dots, n-2, \quad \alpha^{(n-1)} M_0 = 0.$$

PROOF: In the sequel, for any vector $\alpha \in \mathbb{F}^k$ let α_l denote the l -th coordinate of α . By assumption

$$\{\alpha \in \mathbb{F}^k \mid \alpha_l = 0 \text{ for } l \in \mathcal{L}_i\} \subseteq \ker M_i \text{ for all } i \in \mathcal{I}. \quad (3.7)$$

Let $\alpha^{(0)} \in \mathbb{F}^k$ such that $\alpha^{(0)} M_0 = v$. By assumption on the rows of M_0 we may assume without loss of generality

$$\alpha_l^{(0)} = 0 \text{ for } l \notin \mathcal{L}_0. \quad (3.8)$$

Define $\alpha^{(1)} \in \mathbb{F}^k$ via

$$\alpha_l^{(1)} = \begin{cases} \alpha_l^{(0)}, & \text{if } l \in \mathcal{L}_1 \cap \mathcal{L}_0 \\ 0, & \text{if } l \notin \mathcal{L}_1 \cap \mathcal{L}_0. \end{cases} \quad (3.9)$$

Then (3.8) and (3.9) show that $\alpha_l^{(1)} - \alpha_l^{(0)} = 0$ for $l \in \mathcal{L}_1$ and thus $\alpha^{(1)} - \alpha^{(0)} \in \ker M_1$. Assume now we proceed in this way and have constructed $\alpha^{(j)} \in \mathbb{F}^k$ such that

$$\alpha_l^{(j)} = \begin{cases} \alpha_l^{(j-1)}, & \text{if } l \in \bigcap_{i=0}^j \mathcal{L}_i \\ 0, & \text{if } l \notin \bigcap_{i=0}^j \mathcal{L}_i \end{cases} \quad (3.10)$$

and such that $\alpha^{(j)} - \alpha^{(j-1)} \in \ker M_j$. Then put

$$\alpha_l^{(j+1)} = \begin{cases} \alpha_l^{(j)}, & \text{if } l \in \bigcap_{i=0}^{j+1} \mathcal{L}_i \\ 0, & \text{if } l \notin \bigcap_{i=0}^{j+1} \mathcal{L}_i. \end{cases} \quad (3.11)$$

Then $(\alpha^{(j+1)} - \alpha^{(j)})_l = 0$ for all $l \in \bigcap_{i=0}^{j+1} \mathcal{L}_i$ by (3.11) and $\alpha_l^{(j+1)} = \alpha_l^{(j)} = 0$ for all $l \in \mathcal{L}_{j+1} \setminus \bigcap_{i=0}^j \mathcal{L}_i$ by (3.10) and (3.11). Hence $(\alpha^{(j+1)} - \alpha^{(j)})_l = 0$ for all $l \in \mathcal{L}_{j+1}$ and thus $\alpha^{(j+1)} - \alpha^{(j)} \in \ker M_{j+1}$. Thus, proceeding with this construction we find $\alpha^{(n-1)} \in \mathbb{F}^k$ such that $\alpha^{(n-1)} - \alpha^{(n-2)} \in \ker M_{n-1}$ and

$$\alpha_l^{(n-1)} = 0 \text{ for } l \notin \bigcap_{i=0}^{n-1} \mathcal{L}_i.$$

But then $\alpha^{(n-1)} = 0$ because $\bigcap_{i=0}^{n-1} \mathcal{L}_i = \{l \in k \mid i \in (a_l, b_l] \text{ for all } i \in \mathcal{I}\} = \emptyset$, where the latter is due to the fact that $(a, b] \neq \mathcal{I}$ by definition of half-open intervals in (3.1). Hence $\alpha^{(n-1)} M_n = 0$ as desired.

Observe that the proof actually shows that there is even a path from $v \in V_0$ to $0 \in V_{n-1}$ because $\alpha^{(n-1)} = 0$. \square

A main result of [9] is the construction of a certain characteristic matrix for a given code from which all minimal trellises can be derived. We will define the characteristic matrix slightly differently, namely based on certain properties rather than on the outcome of a particular procedure. This will facilitate later considerations. The existence of this object will then follow from the algorithm derived in [9]. However, some care has to be exercised due to the fact that our characteristic matrix will not be uniquely determined by the code (see Example 3.12 below).

Definition 3.7 Let $\mathcal{C} \subseteq \mathbb{F}^n$ be a k -dimensional code with support \mathcal{I} . A *characteristic pair* of \mathcal{C} is defined to be a pair (X, \mathcal{T}) , where

$$X = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \in \mathbb{F}^{n \times n} \text{ and } \mathcal{T} = [(a_l, b_l], l = 1, \dots, n] \quad (3.12)$$

have the following properties

- (i) $\text{im } X = \mathcal{C}$, that is, $\{x_1, \dots, x_n\}$ forms a generating set of \mathcal{C} .
- (ii) $(a_l, b_l]$ is a span of x_l for $l = 1, \dots, n$.
- (iii) a_1, \dots, a_n are distinct and b_1, \dots, b_n are distinct.
- (iv) For all $j \in \mathcal{I}$ there exist exactly $n - k$ row indices l_1, \dots, l_{n-k} such that $j \in (a_{l_i}, b_{l_i}]$ for $i = 1, \dots, n - k$.

We call X a *characteristic matrix* of \mathcal{C} and \mathcal{T} the *characteristic span list*. The *span matrix* of the characteristic pair (X, \mathcal{T}) is defined to be the matrix

$$S = (s_{lj})_{\substack{l=1, \dots, n \\ j=0, \dots, n-1}} \in \mathbb{N}_0^{n \times n}, \text{ where } s_{lj} = \begin{cases} 1, & \text{if } j \in (a_l, b_l], \\ 0, & \text{if } j \notin (a_l, b_l]. \end{cases}$$

As we will see below, the characteristic span list is, up to ordering, uniquely determined by the code \mathcal{C} . This is one of the results proven in [9]. The only difference between our definition and the one of Koetter/Vardy is that in [9] for each span $(a_l, b_l]$ the corresponding row in X is the unique lexicographically first codeword having this span (with a slight modification in the non-binary case). As Koetter/Vardy point out in [9, Remark, p. 2095] there is no need for this particular choice.

Here are some first simple properties of characteristic pairs.

Remark 3.8 (a) Every column of the span matrix S has exactly $n - k$ entries equal to 1 and k entries equal to 0. This is simply a reformulation of Definition 3.7(iv).

(b) Let σ be the left shift on \mathbb{F}^n as in Remark 3.4. If (X, \mathcal{T}) as in (3.12) is a characteristic pair for \mathcal{C} , then (X^*, \mathcal{T}^*) is a characteristic pair of $\sigma(\mathcal{C})$, where

$$X^* := \begin{pmatrix} \sigma(x_1) \\ \vdots \\ \sigma(x_n) \end{pmatrix} \text{ and } \mathcal{T}^* := [(a_l - 1, b_l - 1], l = 1, \dots, n].$$

(c) By part (iv) of the definition there exist exactly k linear spans in the span list \mathcal{T} (recall that $(a, b]$ is linear iff $0 \notin (a, b]$). Therefore and due to (iii), the corresponding rows of X form an MSGM of \mathcal{C} in the sense of [12, Def. 6.2, Thm. 6.11]. Likewise, the previous item shows that for each $j = 0, \dots, n - 1$ a characteristic matrix contains a shifted MSGM for the shifted code $\sigma^j(\mathcal{C})$. In this sense, the rows of a characteristic matrix may be called a set of “shortest generators” of \mathcal{C} (and all its cyclic shifts) in the sense of [3].

So far we have not established the existence of characteristic pairs for a given code. This crucial result, along with the uniqueness of the span set, has been proven in [9] and an algorithm has been given for computing a characteristic matrix. All this leads to the following result.

Theorem 3.9 *Let $\mathcal{C} \subseteq \mathbb{F}^n$ be a k -dimensional code with support \mathcal{I} . Then \mathcal{C} has a characteristic pair and the characteristic span list is, up to ordering, uniquely determined by \mathcal{C} . Thus, the rows of the span matrix S in Definition 3.7 are uniquely determined up to ordering.*

PROOF: The existence of a characteristic pair with the properties as in Definition 3.7(i) – (iv) has been shown in [9, Lem. 5.7, Thm. 5.9, Thm. 5.10 and its proof]. Since our notion of characteristic pair is slightly more general than the one in [9], we need to pay special attention to the uniqueness of the characteristic span list. In order to do so let (X, \mathcal{T}) as in (3.12) be a characteristic pair of \mathcal{C} . Fix $j \in \{0, \dots, n - 1\}$ and choose the k indices l_1, \dots, l_k for which $j \notin (a_{l_i}, b_{l_i}]$; they uniquely exist due to property (iv). Then $0 \notin (a_{l_i} - j, b_{l_i} - j]$ and the latter is a span for the shifted row $\sigma^j(x_{l_i})$. Hence those shifted rows have linear spans and, due to property (iii) they form an MSGM of $\sigma^j(\mathcal{C})$ in the sense of [12, Def. 6.2]. By [12, Thm. 6.11] the shifted span list is uniquely determined by the code $\sigma^j(\mathcal{C})$ and hence by \mathcal{C} . All this shows that the span list \mathcal{T} is uniquely determined by \mathcal{C} and coincides with the one of the characteristic matrix as defined in [9, Def. 5.2]. \square

The matrix X of a characteristic pair is, in general, not uniquely determined by \mathcal{C} (up to ordering of the rows), see Example 3.12 below.

The importance of the characteristic pairs is that all minimal trellises of the code can be retrieved from them. This is one of the main results in [9]. Let us first introduce the following notion.

Definition 3.10 Let $\mathcal{C} \subseteq \mathbb{F}^n$ be a k -dimensional code with support \mathcal{I} and let (X, \mathcal{T}) be a characteristic pair of \mathcal{C} as in (3.12). Any trellis of the form $T_{x_{l_1}, (a_{l_1}, b_{l_1}]} \times \dots \times T_{x_{l_k}, (a_{l_k}, b_{l_k}]}$, where x_{l_1}, \dots, x_{l_k} are linearly independent rows of X , is called a $KV_{(X, \mathcal{T})}$ -trellis of \mathcal{C} . Every trellis that is a $KV_{(X, \mathcal{T})}$ -trellis for some characteristic pair (X, \mathcal{T}) of \mathcal{C} is called a KV -trellis of \mathcal{C} .

Remark 3.11 By Theorem 3.3 every KV-trellis is one-to-one. Moreover, if $T = T_{x_{l_1},(a_{l_1},b_{l_1})} \times \dots \times T_{x_{l_k},(a_{l_k},b_{l_k})}$ is a $\text{KV}_{(X,\mathcal{T})}$ -trellis, then the SCP of T is given by $(v_1, \dots, v_n)S$, where S is the span matrix of the characteristic pair (X, \mathcal{T}) and $v_i = 1$ if $i \in \{l_1, \dots, l_k\}$ and $v_i = 0$ else. This follows immediately from Theorem 3.3. As a consequence, the SCP (s_0, \dots, s_{n-1}) of a KV-trellis satisfies $s_i \leq \min\{k, n-k\}$ for all $i \in \mathcal{I}$, which generalizes the well-known Wolf bound [20] (see also [18, Thm. 5.5]) to tail-biting trellises. Finally, since a KV-trellis is a product trellis where the starting points of the spans are distinct and so are the end points, the formulas for the ECP and SCP given in Proposition 3.5 apply.

Unfortunately, the set of $\text{KV}_{(X,\mathcal{T})}$ -trellises depends on the choice of the characteristic pair (X, \mathcal{T}) . Indeed, we have the following example.

Example 3.12 Let $\mathbb{F} = \mathbb{F}_2$ and $\mathcal{C} = \{(0000), (1100), (0111), (1011)\} \subseteq \mathbb{F}_2^4$. The two pairs (X, \mathcal{T}) and (X', \mathcal{T}) , where

$$X = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & 1 \end{pmatrix}, \quad X' = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 \\ 1 & 0 & 1 & 1 \end{pmatrix}, \quad \mathcal{T} = [(0, 1], (1, 3], (2, 0], (3, 2)], \quad (3.13)$$

are both characteristic pairs of \mathcal{C} . Their common span matrix is

$$S = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{pmatrix}.$$

The last two rows of X are linearly independent and lead to a $\text{KV}_{(X,\mathcal{T})}$ -trellis with SCP $(2, 1, 1, 1)$; see Remark 3.11. But since the last two rows of X' are linearly dependent, this SCP does not appear for any $\text{KV}_{(X',\mathcal{T})}$ -trellis of \mathcal{C} as one can see directly from the span matrix S . It is easy to see that the two matrices X and X' are the only two characteristic matrices for the code \mathcal{C} , up to ordering of the rows. Indeed, examining the set of codewords it is clear that for the spans $(0, 1]$, $(1, 3]$, and $(2, 0]$ there is exactly one codeword having this span, while the span $(3, 2]$ is attained by the two codewords (0111) and (1011) . This leads exactly to the two choices of characteristic matrices given in (3.13). Since (0111) is the lexicographically first of those two vectors, the matrix X is the characteristic matrix singled out in [9].

The following result expresses the fact that the list of spans uniquely determines the structure of KV-trellises. For structural isomorphisms recall Definition 2.1(e).

Corollary 3.13 Let $T = T_{G,S}$ and $T' = T_{G',S'}$ be KV-trellises of \mathcal{C} . Then the following are equivalent:

- (i) T and T' are structurally isomorphic.
- (ii) T and T' have the same SCP and ECP.
- (iii) $S = S'$, up to ordering.

PROOF: (i) \Rightarrow (ii) is obvious.

(ii) \Rightarrow (iii) From Proposition 3.5(a) it follows that the sets \mathcal{A} and \mathcal{A}' of starting points of the spans in S and S' , respectively, are identical. But for KV-trellises any starting point uniquely

determines its span, see Theorem 3.9. Hence $\mathcal{S} = \mathcal{S}'$, up to ordering.

(iii) \Rightarrow (i) Without loss of generality, let us assume $\mathcal{S} = \mathcal{S}'$. Then $M_i = M'_i$ for the product trellis matrices of T and T' , respectively, defined in (3.4). But then the edge spaces of T and T' are given by $E_i = \text{im}(M_i, G_i, M_{i+1})$ and $E'_i = \text{im}(M_i, G'_i, M_{i+1})$, where the i -th column of G and G' is denoted by G_i and G'_i , respectively. By Proposition 3.5 we have $\dim E_i = \dim E'_i$. This shows that $\phi_i(\alpha(M_i, G_i, M_{i+1})) = \alpha(M_i, G'_i, M_{i+1})$ is a well-defined structural isomorphism between T and T' . \square

It is worth observing that the implication (ii) \Rightarrow (iii) is not true in general for product trellises. For instance, the two matrices

$$G = \begin{pmatrix} 1 & 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 1 \\ 1 & 1 & 1 & 0 & 1 \end{pmatrix}, \quad G' = \begin{pmatrix} 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 & 1 \end{pmatrix} \in \mathbb{F}_2^{3 \times 5}$$

generate the same code of dimension 3. Choosing the span list $\mathcal{S} = [(3, 1), (0, 4), (4, 2)]$ for G and $\mathcal{S}' = [(4, 1), (3, 2), (0, 4)]$ for G' results in product trellises $T_{G, \mathcal{S}}$ and $T_{G', \mathcal{S}'}$ with the same SCP $(2, 3, 2, 1, 2)$ and ECP $(3, 3, 2, 2, 3)$ even though $\mathcal{S} \neq \mathcal{S}'$. However, one can check that the two trellises are not structurally isomorphic. As a consequence, the question whether (i) above is equivalent to (iii) remains open for (one-to-one) product trellises.

Let us now state the main result of [9].

Theorem 3.14 ([9, Thm. 5.5]) *Let $\mathcal{C} \subseteq \mathbb{F}^n$ be a k -dimensional code with support \mathcal{I} and let T be a minimal trellis of \mathcal{C} in the sense of Definition 2.2(a). Then T is one-to-one and there exists a characteristic pair (X, \mathcal{T}) such that T is a $\text{KV}_{(X, \mathcal{T})}$ -trellis.*

PROOF: Since our result is phrased differently from what is in [9, Thm. 5.5] we think it is worthwhile to give a proof. First of all, T is a product trellis by [9, Thm. 4.2] and is one-to-one due to [9, Lemma 5.2]. Hence T is a product trellis as in our Theorem 3.3 because, due to minimality, the trellis is non-mergeable and thus we may assume shortest spans for all generators; see also the paragraphs following Theorem 3.3. Thus, $T = T_{x_1, (a_1, b_1)} \times \dots \times T_{x_k, (a_k, b_k)}$ for certain $x_l \in \mathcal{C}$ with spans (a_l, b_l) . Now [9, Lemma 5.4] implies that $(x_1, (a_1, b_1)), \dots, (x_k, (a_k, b_k))$ are characteristic generators in the sense of [9, Def 5.2]. But this simply means that this list can be extended to a characteristic pair (X, \mathcal{T}) of \mathcal{C} in the sense of our Definition 3.7 and thus T is a $\text{KV}_{(X, \mathcal{T})}$ -trellis. \square

One should bear in mind that not all KV-trellises are minimal. For instance, the last two rows of X in Example 3.12 are linearly independent and give rise to a trellis with SCP $(2, 1, 1, 1)$, but the first two rows of X are also linearly independent and lead to a trellis with SCP $(0, 1, 1, 1)$. This shows that the first trellis is not minimal in the sense of Definition 2.2(a). A larger example is given in [9, Exa. 6.1].

It is important to notice that the above result does not imply that all minimal trellises of \mathcal{C} arise from a fixed characteristic pair. Indeed, we have the following example.

Example 3.15 Let $\mathbb{F} = \mathbb{F}_2$ and $\mathcal{C} := \text{im } G$, where

$$G = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 1 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 \end{pmatrix} \in \mathbb{F}_2^{3 \times 6}.$$

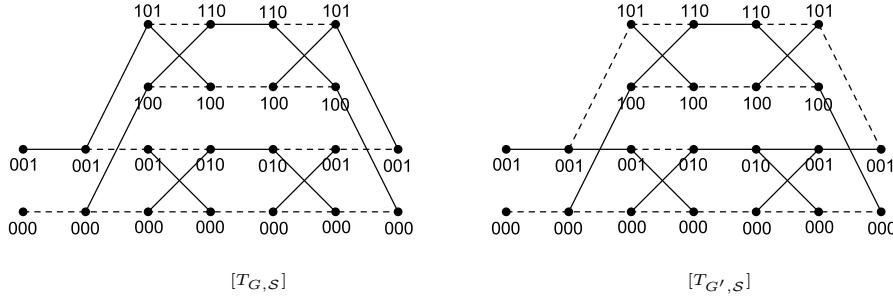
Put

$$X = \begin{pmatrix} 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 \end{pmatrix}, \quad X' = \begin{pmatrix} 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 0 & 1 & 1 \end{pmatrix}$$

and $\mathcal{T} = [(0, 3], (1, 5], (2, 4], (3, 0], (5, 1], (4, 2]]$. It is easy to verify that (X, \mathcal{T}) and (X', \mathcal{T}) are both characteristic pairs of \mathcal{C} . Put $\mathcal{S} = [(1, 5], (2, 4], (4, 2]]$. Then the product trellises $T = T_{G, \mathcal{S}}$ and $T' = T_{G', \mathcal{S}}$, where

$$G' = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 0 & 1 & 1 \end{pmatrix}$$

are both KV-trellises based on the same choice of characteristic spans; the first one is a $\text{KV}_{(X, \mathcal{T})}$ -trellis and the second one is the corresponding $\text{KV}_{(X', \mathcal{T})}$ -trellis. The KV-trellises are minimal as one can check by going through the SCP's of all possible $\text{KV}_{(X, \mathcal{T})}$ -trellises. (There are 10 choices for selecting 3 linearly independent rows in X and none of them leads to a smaller trellis.) As one can verify directly from the graphs below, the trellises T and T' are structurally isomorphic, but not isomorphic.



In light of the previous example it should be observed that, at this point, the following question is open. Given a code $\mathcal{C} \subseteq \mathbb{F}^n$ and fix a characteristic pair (X, \mathcal{T}) . Is then each minimal trellis of \mathcal{C} structurally isomorphic to a $\text{KV}_{(X, \mathcal{T})}$ -trellis? From Corollary 3.13 we know that the SCP and ECP of the given trellis uniquely specify the row selection of X that may result in a potentially structurally isomorphic $\text{KV}_{(X, \mathcal{T})}$ -trellis T' . But it is not clear whether this row selection is linearly independent. Only then the trellis T' represents \mathcal{C} and thus is a $\text{KV}_{(X, \mathcal{T})}$ -trellis.

Remark 3.16 As shown in [9, Thm. 5.6], the same result as in Theorem 3.14 applies to many other notions of minimality for tail-biting trellises, see [9, p. 2085].

4 The Tail-Biting BCJR-Construction

In this section we will present a different way of producing tail-biting trellises for a given code. These BCJR-constructions were introduced in [15]. We will show that the resulting trellises are non-mergeable and that product trellises can always be merged into a corresponding BCJR-trellis by taking suitable quotients. We will see that KV-trellises are isomorphic to

their BCJR-counterpart, and, consequently, non-mergeable. Throughout let \mathcal{C} be as in (2.1) and (2.2).

Theorem 4.1 ([15, Lemma 2]) *Let $D \in \mathbb{F}^{k \times (n-k)}$ be any matrix and for $i \in \mathcal{I}$ define the matrices*

$$N_i = \begin{cases} D, & \text{if } i = 0 \\ N_{i-1} + G_{i-1}H_{i-1}, & \text{if } i > 0 \end{cases} \quad (4.1)$$

Then $N_n = N_0$ and the trellis $T := T_{(G,H,D)}$ with vertex spaces $V_i := \text{im } N_i \subseteq \mathbb{F}^{n-k}$ and edge spaces $E_i = \text{im}(N_i, G_i, N_{i+1})$ is a linear, reduced and biproper trellis representing the code \mathcal{C} , that is, $\mathcal{C}(T) = \mathcal{C}$. We call D the displacement matrix for the trellis T .

The fact $N_n = N_0$ simply follows from the identity $N_i = N_0 + \sum_{j=0}^{i-1} G_j H_j$, $i = 1, \dots, n$ along with $0 = GH^\top = \sum_{j=0}^{n-1} G_j H_j$. The reducedness and biproperness are a consequence of the very definition of the vertex and edge spaces along with the recursive definition of the matrices N_i . The displacement matrix D can be interpreted as the (circular) past at time zero of each cycle in $T_{(G,H,D)}$. If $D = 0$, each cycle starts with zero history and the trellis $T_{(G,H,D)}$ is conventional. In fact, it is simply the classical BCJR-trellis and thus a minimal trellis, see [12, Sec. IV] and [1]. This case will also show up as a special version of Definition 4.5 below, in which we define what will be called a tail-biting BCJR-trellis. One should regard the displacement matrix D as a design parameter for the construction of good trellises. In Definition 4.5 we will restrict ourselves to particularly useful choices of this parameter, which then will always lead to non-mergeable trellises.

Remark 4.2 For every $U \in GL_k(\mathbb{F})$ the trellises $T_{(G,H,D)}$ and $T_{(UG,H,UD)}$ are identical and for all $V \in GL_{n-k}(\mathbb{F})$ the trellises $T_{(G,H,D)}$ and $T_{(G,VH,DV^\top)}$ are isomorphic and the isomorphisms on the vertex spaces are given by the restrictions of the linear map V^\top to those spaces.

The statement $\mathcal{C}(T) = \mathcal{C}$ of Theorem 4.1 also follows from the following result which will be helpful later on.

Proposition 4.3 *Let $T = T_{(G,H,D)}$ be as in Theorem 4.1. Furthermore, let $\alpha^{(i)} \in \mathbb{F}^k$ be such that the sequence of edges $(\alpha^{(i)}N_i, \alpha^{(i)}G_i, \alpha^{(i)}N_{i+1})$, $i \in \mathcal{I}$, is a path through T , that is, $\alpha^{(i)}N_{i+1} = \alpha^{(i+1)}N_{i+1}$ for $i = 0, \dots, n-2$. Then the edge-label sequence of the path is a codeword if and only if the path is a cycle. In other words,*

$$(\alpha^{(0)}G_0, \dots, \alpha^{(n-1)}G_{n-1}) \in \mathcal{C} \iff \alpha^{(n-1)}N_0 = \alpha^{(0)}N_0.$$

PROOF: From the assumptions as well as the definition of the matrices N_i we obtain $\alpha^{(1)}N_1 = \alpha^{(0)}N_1 = \alpha^{(0)}N_0 + \alpha^{(0)}G_0H_0$ and inductively $\alpha^{(i+1)}N_{i+1} = \alpha^{(0)}N_0 + \sum_{j=0}^i \alpha^{(j)}G_jH_j$, for $i = 0, \dots, n-2$. Hence $\alpha^{(n-1)}N_{n-1} = \alpha^{(0)}N_0 + \sum_{j=0}^{n-2} \alpha^{(j)}G_jH_j$ and this implies

$$\alpha^{(n-1)}N_0 = \alpha^{(n-1)}N_{n-1} + \alpha^{(n-1)}G_{n-1}H_{n-1} = \alpha^{(0)}N_0 + \sum_{j=0}^{n-1} \alpha^{(j)}G_jH_j$$

This leads to $\alpha^{(n-1)}N_0 = \alpha^{(0)}N_0 \iff \sum_{i=0}^{n-1} \alpha^{(i)}G_iH_i = 0$ and the latter is the case if and only if $(\alpha^{(0)}G_0, \dots, \alpha^{(n-1)}G_{n-1}) \in \ker H^\top = \mathcal{C}$, as desired. \square

One of the nice properties of the trellis construction in Theorem 4.1 is that it respects duality. Indeed, the following is straightforward; see also [15, Lemma 12].

Proposition 4.4 Let $T = T_{(G,H,D)}$ be as in Theorem 4.1. Then $\hat{T} := T_{(H,G,D^\top)}$ is a trellis representing the dual code \mathcal{C}^\perp . Moreover, the vertex spaces of \hat{T} are given by $\text{im } N_i^\top$, $i \in \mathcal{I}$, and thus the trellises T and \hat{T} have the same SCP, given by (s_0, \dots, s_{n-1}) , where $s_i = \text{rk } N_i$. As a consequence, if T is a minimal trellis for \mathcal{C} , then \hat{T} is a minimal trellis for \mathcal{C}^\perp .

Later on in the proof of Theorem 5.3 we will see that in the special case where $T_{(G,H,D)}$ is a minimal KV-trellis, this trellis dualization coincides with the construction given by Forney in [5, Sec. VII-D and Thm. 8.4] based on dualizing the edge spaces (local constraints) with respect to a particular inner product.

A particular instance of the trellis $T_{(G,H,D)}$ is obtained by choosing the displacement matrix based on given spans of the generator matrix G . This will relate this construction to the product construction of the previous section.

Definition 4.5 Let $\mathcal{S} := [(a_l, b_l], l = 1, \dots, k]$ be a span list of G . Then the trellis $T_{(G,H,\mathcal{S})}$ defined as $T_{(G,H,D)}$, where

$$D = \begin{pmatrix} d_1 \\ \vdots \\ d_k \end{pmatrix} \in \mathbb{F}^{k \times (n-k)} \text{ with } d_l = \sum_{j=a_l}^{n-1} g_{lj} H_j, \quad (4.2)$$

is called a *(tail-biting) BCJR-trellis* of \mathcal{C} .

It should be noted that we define a BCJR-trellis more restrictive than Nori/Shankar in [15]: while we use this terminology only for trellises $T_{(G,H,D)}$, where the displacement matrix is based on a given span list as in (4.2), Nori/Shankar use the term BCJR-trellis for all displacement matrices.

Remark 4.6 From Remark 4.2 it follows that $T_{(G,H,\mathcal{S})}$ is isomorphic to $T_{(G,VH,\mathcal{S})}$ for all $V \in GL_{n-k}(\mathbb{F})$. Indeed, if D is as in (4.2) then DV^\top is the according displacement matrix for $T_{(G,VH,\mathcal{S})}$. As a consequence, V^\top restricted to all vertex spaces furnishes an isomorphism between the two trellises. However, if G and $UG \in \mathbb{F}^{k \times n}$, $U \in GL_k(\mathbb{F})$, are two generator matrices with the same span list \mathcal{S} , then the BCJR-trellises $T_{(G,H,\mathcal{S})}$ and $T_{(UG,H,\mathcal{S})}$ need not be isomorphic. This can be seen from Example 3.15. Indeed, the two (non-isomorphic) trellises given in that example are BCJR-trellises, which will follow from Theorem 4.12 presented below.

Notice that if all spans are linear, then D is the zero matrix (see also Proposition 4.8 below), and the resulting trellis is the conventional BCJR-trellis. It is well-known that the conventional BCJR-trellis is minimal, see [12, Sec. IV], and thus non-mergeable. Later on we will see that in the tail-biting case, BCJR-trellises are non-mergeable as well, but they are not necessarily minimal.

Just like the product trellises in Remark 3.4 the BCJR-construction is compatible with the shift.

Remark 4.7 Let \mathcal{S} and D be as in Definition 4.5. As in Remark 3.4 let $G^* \in \mathbb{F}^{k \times n}$ be the matrix consisting of the shifted rows $\sigma(g_l)$, $l = 1, \dots, k$ and define H^* accordingly. Then $\mathcal{S}^* := [(a_l - 1, b_l - 1], l = 1, \dots, k]$ is a span list for the rows of G^* . Let N_i , $i \in \mathcal{I}$, be the vertex matrices as in (4.1) for the BCJR-trellis $T_{(G,H,\mathcal{S})}$. Then the vertex matrices for the

BCJR-trellis $T_{(G^*, H^*, S^*)}$ are given by $N_i^* = N_{i+1}$ for $i \in \mathcal{I}$. Indeed, by (4.2) the l -th row of N_0^* is defined as

$$\sum_{j=a_l-1}^{n-1} g_{lj}^* H_j^* = \sum_{j=a_l-1}^{n-1} g_{l,j+1} H_{j+1} = \sum_{j=a_l}^{n-1} g_{lj} H_j + g_{l0} H_0$$

and the latter is the l -th row of N_1 . Now the rest follows inductively due to (4.1).

In the sequel we will show that if the pair (G, S) originates from a characteristic pair of \mathcal{C} , then the trellis $T_{(G, H, S)}$ is isomorphic to the corresponding KV-trellis $T_{G, S}$.

We begin by showing the following.

Proposition 4.8 *Let $T_{(G, H, S)}$ be as in Definition 4.5 and let N_i , $i \in \mathcal{I}$, be defined as in Theorem 4.1. Then*

- (a) *If $(a_l, b_l]$ is a linear span, then $d_l = 0$.*
- (b) *If $i \in \mathcal{I}$ and $l \in \{1, \dots, k\}$ are such that $i \notin (a_l, b_l]$, then the l -th row of N_i is zero.*

PROOF: (a) If $(a_l, b_l]$ is linear then all nonzero entries of g_l are in the interval $[a_l, n-1]$ and therefore $d_l = \sum_{j=a_l}^{n-1} g_{lj} H_j = \sum_{j=0}^{n-1} g_{lj} H_j = 0$ due to $GH^\top = 0$.

(b) Recall that $N_0 = D$ as well as the fact that $(a_l, b_l]$ is a linear span iff $0 \notin (a_l, b_l]$. Therefore, the case $i = 0$ has been proven in (a). Thus, let $i \geq 1$. Then, by definition, $N_i = D + \sum_{j=0}^{i-1} G_j H_j$.

Suppose first that the span $(a_l, b_l]$ is circular, hence in particular, non-empty. Then $i \notin (a_l, b_l]$ is equivalent to $i \in (b_l, a_l]$, which is linear; see also (3.2). Thus, the l -th row of N_i is

$$d_l + \sum_{j=0}^{i-1} g_{lj} H_j = \sum_{j=a_l}^{n-1} g_{lj} H_j + \sum_{j=0}^{i-1} g_{lj} H_j = \sum_{j=0}^{n-1} g_{lj} H_j = 0,$$

where the second identity holds true because $g_{lj} = 0$ for $j \in [i, a_l - 1]$.

Let now the span $(a_l, b_l]$ be linear. In this case $d_l = 0$ by part (a) and the l -th row of N_i is given by

$$\sum_{j=0}^{i-1} g_{lj} H_j. \tag{4.3}$$

Now $i \notin (a_l, b_l]$ amounts to either $i \in [0, a_l]$ or $i \in (b_l, n-1]$. In the first case, the interval $[0, i-1]$ and the closed span $[a_l, b_l]$ of g_l are disjoint while in the second case the interval $[0, i-1]$ contains $[a_l, b_l]$. Hence in both cases the sum in (4.3) is zero. \square

Now we are in a position to prove the non-mergeability of the BCJR-trellises. This result generalizes the well-known fact that the conventional BCJR-trellis is minimal (thus non-mergeable), see [12, Sec. V]. It is worth noticing that, just like in the conventional case, the corollary is true even without assuming that the starting points a_1, \dots, a_k of the spans (or the endpoints b_1, \dots, b_k) are distinct.

Corollary 4.9 *Any BCJR-trellis $T := T_{(G, H, S)}$ is non-mergeable.*

PROOF: Let V_i be the i -th vertex space of $T_{(G, H, S)}$. Let us first suppose that the vertices $v, v' \in V_0$ are mergeable. Then the edge-label sequence of each path through T starting at $v \in V_0$ and ending in $v' \in V_0$ must be a codeword in \mathcal{C} (because after merging this path

would be a cycle). But then by linearity of the trellis the edge-label sequence of each path through T starting at $v - v' \in V_0$ and ending in $0 \in V_0$ must be a codeword in \mathcal{C} as well. Due to Proposition 4.8(b) we may apply Lemma 3.6 and conclude that such a path does indeed exist in T . But then Proposition 4.3 implies that this path is a cycle. Hence $v - v' = 0$. All this shows that no vertices in V_0 are mergeable. Now the non-mergeability of vertices in any V_i follows from the shift property of T as described in Remark 4.7. \square

It should be noted that if the displacement matrix D is not as in Definition 4.5, then the trellises $T_{(G,H,D)}$ as in Theorem 4.1 are, in general, mergeable even if the starting points (resp., end points) of the spans are distinct. An example is given in [15, Exa. 9, Fig. 14]. One can easily see how to merge that trellis. Moreover, that trellis has SCP (2, 2, 2, 2) and therefore Proposition 4.8(b) implies that the underlying displacement matrix is not as in Definition 4.5 and thus the trellis is not a BCJR-trellis in our sense. Furthermore, that trellis does not contain any paths connecting a nonzero vertex in V_0 with $0 \in V_0$ and therefore the main argument of the proof above does not apply.

At this point it should be mentioned that BCJR-trellises are in general not one-to-one, not even if the starting points (resp. endpoints) of the spans are distinct. This is illustrated by the following example taken from [15, Exa. 8].

Example 4.10 Let $\mathcal{C} = \text{im } G = \ker H^T$, where

$$G = \begin{pmatrix} 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix} \in \mathbb{F}_2^{2 \times 3}, \text{ and } \mathcal{S} = [(0, 2), (1, 0)], H = (1 \ 1 \ 1).$$

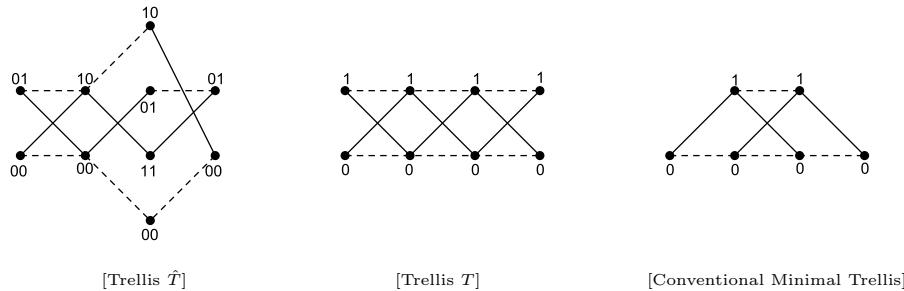
The BCJR-trellis $T := T_{(G,H,\mathcal{S})}$ has edge spaces $E_i = \text{im}(N_i, G_i, N_{i+1})$ given by

$$E_0 = \text{im} \begin{pmatrix} 0 & 1 & 1 \\ 1 & 1 & 0 \end{pmatrix}, \quad E_1 = \text{im} \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \end{pmatrix}, \quad E_2 = \text{im} \begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix}.$$

This shows that $1 \xrightarrow{0} 1 \xrightarrow{0} 1 \xrightarrow{0} 1$ is a non-trivial cycle in T with the zero codeword as edge-label sequence; see also the figure below. Hence T is not one-to-one. Let us also consider the associated product trellis $\hat{T} = T_{G,\mathcal{S}}$. In this case the edge spaces are given by

$$\hat{E}_0 = \text{im} \begin{pmatrix} 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 & 0 \end{pmatrix}, \quad \hat{E}_1 = \text{im} \begin{pmatrix} 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 \end{pmatrix}, \quad \hat{E}_2 = \text{im} \begin{pmatrix} 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 \end{pmatrix}.$$

From Theorem 3.3 we know that \hat{T} is one-to-one; see figure below. However, the trellis \hat{T} is mergeable at time $i = 2$ and indeed, merging the states (11), (00) and the states (10), (01) (that is, replacing V_2 by the quotient space $V_2/\text{im}(11)$) leads to the (not one-to-one) trellis $T = T_{(G,H,\mathcal{S})}$. In the following graphs as well as in all later graphs a solid line denotes an edge with label 1, while dashed lines denote the label 0.



The trellis T , being a BCJR-trellis, is non-mergeable. However, it is not minimal as the conventional minimal trellis of \mathcal{C} at the very right shows. It should also be noted that \hat{T} is not (isomorphic to) a KV-trellis. Indeed, as one can easily verify, the unique characteristic pair of \mathcal{C} with span matrix S is given by

$$(X, \mathcal{T}) = \left(\begin{pmatrix} 0 & 1 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix}, [(1, 2), (0, 1), (2, 0)] \right), \quad S = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix}.$$

From this it is obvious that the SCP $(1, 1, 2)$ of \hat{T} cannot be attained by a KV-trellis and neither can the SCP $(1, 1, 1)$ of the trellis T . Moreover, the trellises \hat{T} and T cannot be merged into a KV-trellis of the code.

On the other hand, in Theorem 4.14 we will see that the process of merging a product trellis \hat{T} into the corresponding BCJR-trellis T is always possible (if \hat{T} is mergeable at all) and can be achieved by taking a suitable quotient trellis.

Theorem 4.11 *Let $\mathcal{S} := [(a_l, b_l], l = 1, \dots, k]$ and $T := T_{(G, H, \mathcal{S})}$ be as in Definition 4.5. Let \hat{T} be the product trellis $\hat{T} := T_{G, \mathcal{S}}$. Then the vertex spaces $V_i := \text{im } N_i$ and $\hat{V}_i := \text{im } M_i$, $i \in \mathcal{I}$, of T and \hat{T} , respectively, satisfy*

$$\dim V_i \leq \dim \hat{V}_i \text{ for } i \in \mathcal{I}. \quad (4.4)$$

If equality holds true for all $i \in \mathcal{I}$ in (4.4), that is, T and \hat{T} have the same SCP, then one has the following.

- (a) T and \hat{T} are isomorphic, thus the product trellis \hat{T} is non-mergeable.
- (b) a_1, \dots, a_k are distinct and so are b_1, \dots, b_k . Therefore, the formulas for the ECP and SCP given in Proposition 3.5 apply.

It is worth noting that $T \preceq_{\Theta} \hat{T}$ with respect to the Θ -ordering defined in [9, p. 2084]. While we have seen in Proposition 3.5 that if both a_1, \dots, a_k and b_1, \dots, b_k are distinct, the formulas for the ECP do always apply to product trellises, this is not the case for the BCJR-trellis; see for instance Example 4.10.

PROOF: First of all, by Proposition 4.8(b) and the very definition of the matrices M_i in (3.4) we have

$$\text{col}(N_i) \subseteq \text{col}(M_i), \quad (4.5)$$

where $\text{col}(M) := \{Mw \mid w \in \mathbb{F}^r\}$ denotes the column space of the matrix $M \in \mathbb{F}^{k \times r}$. Hence $\text{rk}(N_i) \leq \text{rk}(M_i)$ and this proves (4.4).

Let us now assume that we have equality in (4.4) for all $i \in \mathcal{I}$.

- (a) The inclusions (4.5) imply

$$\text{col}(N_i) = \text{col}(M_i) \text{ for } i \in \mathcal{I} \quad (4.6)$$

and hence $\ker N_i = \ker M_i$, where, as usual the kernel denotes the left kernel and thus is simply the orthogonal of the column space with respect to the standard inner product on \mathbb{F}^k . As a consequence, the map $\phi_i : \hat{V}_i \longrightarrow V_i$, $\alpha M_i \longmapsto \alpha N_i$ is a well-defined isomorphism. But then the maps ϕ_i give rise to an isomorphism between \hat{T} and T because the edge spaces of \hat{T} are given by $\hat{E}_i = \text{im}(N_i, G_i, N_{i+1})$ while those of T are $E_i = \text{im}(M_i, G_i, M_{i+1})$. The non-mergeability of \hat{T} follows from Corollary 4.9.

(b) Let the index sets \mathcal{L}_i be as in (3.5) again. Then $\text{rk}(M_i) = |\mathcal{L}_i|$ by Proposition 3.5. Since by assumption $\text{rk}(N_i) = \text{rk}(M_i)$ Proposition 4.8(b) implies

the rows of N_i with indices from \mathcal{L}_i are linearly independent. (4.7)

Assume first that for some $l \neq m$ we have $a_l = a_m$. Put $i = a_l + 1 = a_m + 1$. Then the l -th row of N_i is given by

$$\sum_{j=a_l}^{n-1} g_{lj} H_j + \sum_{j=0}^{i-1} g_{lj} H_j = g_{la_l} H_{a_l}$$

and likewise the m -th row of N_i is

$$\sum_{j=a_m}^{n-1} g_{mj} H_j + \sum_{j=0}^{i-1} g_{mj} H_j = g_{ma_l} H_{a_l}. \quad (4.8)$$

Hence these two rows are linearly dependent. Observe that if $a_l \neq b_l$ and $a_m \neq b_m$ then $l, m \in \mathcal{L}_i$ and therefore this linear dependence contradicts (4.7). If $a_l = b_l = a_m \neq b_m$, then $(a_l, b_l) = \emptyset$ and $m \in \mathcal{L}_i$, but $l \notin \mathcal{L}_i$. However, in this case the row g_l is nonzero only at the a_l -th position and this forces $H_{a_l} = 0$. Again, (4.8) contradicts (4.7). The case $b_l \neq a_l = a_m = b_m$ is symmetric and the case $a_l = b_l = a_m = b_m$ cannot occur due to the linear independence of the rows g_l, g_m .

Finally assume that for some $l \neq m$ we have $b_l = b_m$. Then for $i = b_l = b_m$ the l -th row of N_i is $\sum_{j=a_l}^{n-1} g_{lj} H_j + \sum_{j=0}^{i-1} g_{lj} H_j = -g_{lb_l} H_{b_l}$ and likewise the m -th row of N_i is $-g_{mb_l} H_{b_l}$. Now we may argue as in the first case. □

In the proof we used the fact that if equality holds in (4.4) then we have (4.7). With the aid of Proposition 4.8(b) this yields

$$\ker N_i = \{\alpha \in \mathbb{F}^k \mid \alpha_l = 0 \text{ for } l \in \mathcal{L}_i\}. \quad (4.9)$$

Now we are in a position to prove that KV-trellises are BCJR-trellises.

Theorem 4.12 *Let (X, \mathcal{T}) be a characteristic pair of \mathcal{C} as in Definition 3.7. Let $G \in \mathbb{F}^{k \times n}$ consist of k linearly independent rows of X and let \mathcal{S} be the list of corresponding spans taken from \mathcal{T} . Then the trellises $T_{G, \mathcal{S}}$ and $T_{(G, H, \mathcal{S})}$ are isomorphic. As a consequence, KV-trellises are BCJR-trellises and thus non-mergeable.*

PROOF: The last statement follows from Corollary 4.9. For the first part it suffices to prove that $\text{rk}(M_i) = \text{rk}(N_i)$ for all $i \in \mathcal{I}$. Indeed, then the result is immediate with Theorem 4.11. We will proceed in several steps.

1) By Definition 3.7(iv) we may assume without loss of generality that the first k rows of X have linear span (thus $a_l \leq b_l$ for $l = 1, \dots, k$) and the last $n - k$ rows have circular spans. Then those first k rows x_1, \dots, x_k are linearly independent and form an MSGM of \mathcal{C} , see [12, Def. 6.2, Thm. 6.11]. Define the matrix $Z \in \mathbb{F}^{(n-k) \times (n-k)}$ via

$$Z = \begin{pmatrix} z_1 \\ \vdots \\ z_{n-k} \end{pmatrix}, \quad \text{where } z_l = \sum_{j=a_{k+l}}^{n-1} x_{k+l, j} H_j \quad (4.10)$$

(where $x_{k+l} = (x_{k+l, 0}, \dots, x_{k+l, n-1})$ is the $(k+l)$ -th row of X). We claim that Z is nonsingular.

In order to see this, assume $\beta Z = 0$ for some $\beta = (\beta_1, \dots, \beta_{n-k}) \in \mathbb{F}^{n-k}$. Hence

$$\sum_{l=1}^{n-k} \beta_l \sum_{j=a_{k+l}}^{n-1} x_{k+l,j} H_j = 0. \quad (4.11)$$

For $l = 1, \dots, n-k$ define $\hat{x}_{k+l} := (0, \dots, 0, x_{k+l,a_{k+l}}, \dots, x_{k+l,n-1}) \in \mathbb{F}^n$, that is, \hat{x}_{k+l} is the vector with first a_{k+l} entries equal to zero and the last ones identical to those of x_{k+l} . Notice that by definition of the span of a vector $x_{k+l,a_{k+l}} \neq 0$ and remember that a_{k+1}, \dots, a_n are distinct by definition of a characteristic pair. Therefore, $\hat{x}_{k+1}, \dots, \hat{x}_n$ are linearly independent. Moreover, (4.11) implies

$$0 = \sum_{l=1}^{n-k} \beta_l \sum_{j=0}^{n-1} \hat{x}_{k+l,j} H_j = \sum_{j=0}^{n-1} \left(\sum_{l=1}^{n-k} \beta_l \hat{x}_{k+l,j} \right) H_j.$$

But this means that $\tilde{x} \in \ker H^T = \mathcal{C}$, where $\tilde{x} := \sum_{l=1}^{n-k} \beta_l \hat{x}_{k+l}$. If we can show that $\tilde{x} = 0$, then the linear independence of $\hat{x}_{k+1}, \dots, \hat{x}_n$ implies $\beta = 0$, which in turn proves that Z is nonsingular. We proceed by contradiction and assume $\tilde{x} \neq 0$. Then \tilde{x} has a unique linear span, say $(\tilde{a}, \tilde{b}]$, where, by definition of \tilde{x} ,

$$\tilde{a} = \min\{a_{k+l} \mid \beta_l \neq 0\}. \quad (4.12)$$

On the other hand, the first k rows of X form an MSGM of \mathcal{C} and therefore the codeword \tilde{x} is of the form $\tilde{x} = \sum_{l=1}^k \gamma_l x_l$ for some $\gamma_l \in \mathbb{F}$. Now the predictable span property of an MSGM [12, Lem. 6.6] implies that $\tilde{a} = \min\{a_l \mid \gamma_l \neq 0\}$. Along with (4.12) this contradicts the fact that a_1, \dots, a_n are distinct. Thus $\tilde{x} = 0$ as desired.

2) Consider now the matrix D as defined in (4.2). The rows of D corresponding to linear spans are zero (see Proposition 4.8), while the other rows, corresponding to circular spans, are certain rows of the matrix Z in (4.10). Hence the nonsingularity of Z implies that those rows are linearly independent. All this shows that $\text{rk}(D) = \text{rk}(N_0) = |\{l \mid (a_l, b_l) \text{ circular}\}| = |\mathcal{L}_0| = \text{rk}(M_0)$, by virtue of Proposition 3.5.

3) From 2) we have $\text{rk}(N_0) = \text{rk}(M_0)$. Now the identities $\text{rk}(N_i) = \text{rk}(M_i)$ for general $i \in \mathcal{I}$ follow from applying the shift σ and making use of Remarks 3.4, 3.8(b), and 4.7. \square

Theorem 4.12 shows that the set of KV-trellises is a subset of the set of (one-to-one) BCJR-trellises. The following example illustrates that this containment is proper, that is, the set of one-to-one BCJR-trellises is strictly larger than the set of KV-trellises. As a consequence, the class of non-mergeable one-to-one trellises is strictly larger than the class of KV-trellises. At the end of this section we will sketch a proof showing that the class of non-mergeable (one-to-one) trellises coincides with the class of (one-to-one) BCJR-trellises.

Example 4.13 Consider

$$G = \begin{pmatrix} 0 & 1 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 & 1 \end{pmatrix} \in \mathbb{F}_2^{3 \times 5} \text{ with span list } \mathcal{S} = [(1, 3), (3, 0), (2, 1)].$$

Since the rows of G are linearly independent, the product trellis $T := T_{G, \mathcal{S}}$ is one-to-one. Moreover, it has vertex spaces $V_i = \text{im } M_i$, where

$$M_0 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, M_1 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}, M_2 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, M_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}, M_4 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Hence the SCP of T is given by $s = (2, 1, 1, 2, 2)$. One can straightforwardly verify that a characteristic pair of the code $\mathcal{C} = \text{im } G \subseteq \mathbb{F}_2^5$ is given by (X, \mathcal{T}) with span matrix S , where

$$X = \begin{pmatrix} 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 1 & 1 & 0 \end{pmatrix}, \quad \mathcal{T} = [(0, 2], (1, 3], (3, 4], (4, 0], (2, 1)], \quad S = \begin{pmatrix} 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 1 \end{pmatrix}.$$

Using Remark 3.11 one can see directly that no three rows of X and their according spans will lead to the SCP above. Hence T is not isomorphic to a KV-trellis. Let us use the parity check matrix

$$H = \begin{pmatrix} 1 & 0 & 1 & 1 & 1 \\ 0 & 1 & 1 & 0 & 0 \end{pmatrix}$$

in order to compute the BCJR-trellis $T_{(G, H, S)}$ of \mathcal{C} . Applying (4.2) and (4.1) leads to

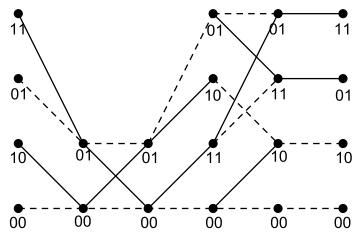
$$N_0 = \begin{pmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad N_1 = \begin{pmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 1 \end{pmatrix}, \quad N_2 = \begin{pmatrix} 0 & 1 \\ 0 & 0 \\ 0 & 0 \end{pmatrix}, \quad N_3 = \begin{pmatrix} 1 & 0 \\ 0 & 0 \\ 1 & 1 \end{pmatrix}, \quad N_4 = \begin{pmatrix} 0 & 0 \\ 1 & 0 \\ 1 & 1 \end{pmatrix}.$$

Hence $\text{col}(N_i) = \text{col}(M_i)$ for all $i \in \{0, \dots, 4\}$ and due to Theorem 4.11 the trellis T is isomorphic to the BCJR-trellis $T_{(G, H, S)}$ and, as a consequence, T is non-mergeable. This can also be seen directly from the concrete trellis $T_{(G, H, S)}$ in the figure below. Let us also briefly verify the statements about the ECP as given in Proposition 3.5. The edge spaces E_i are given by $E_i = \text{im}(N_i, G_i, N_{i+1})$, where G_i is the i -th column of G . For sake of space we display the matrix $(N_0|G_0|N_1| \dots |N_4|G_4|N_0)$. It is given by

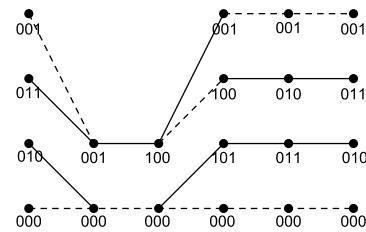
$$\left(\begin{array}{c|c|c|c|c|c|c|c|c|c|c|c|c|c|c|c} 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 & 0 & 1 & 1 & 1 & 0 & 1 \end{array} \right)$$

and shows that the ECP is given by $e = (2, 2, 2, 3, 2)$. Hence $e = s + (0, 1, 1, 1, 0)$ and the latter is indeed the indicator function of the set $\mathcal{A} = \{1, 2, 3\}$ of starting points of the spans. Likewise, we have $e = (1, 1, 2, 2, 2) + (1, 1, 0, 1, 0)$ and the latter is the indicator function of $\mathcal{B} = \{0, 1, 3\}$. All this illustrates the formulas given in Proposition 3.5.

Finally, from the graph below it can easily be checked that the dual trellis $T_{(H, G, N_0^\top)}$, representing $\mathcal{C}^\perp = \text{im } H$ (see Proposition 4.4), is mergeable at time $i = 0$. As a consequence, the dual of a BCJR-trellis need not be a BCJR-trellis. Even more, the dual graph is disconnected and there does not exist a path from $(011) \in \text{im } N_0^\top$ to $(000) \in \text{im } N_0^\top$. Hence the dual trellis does not satisfy Lemma 3.6.



[Trellis $T_{(G, H, S)}$]



[Trellis $T_{(H, G, N_0^\top)}$]

In [15, Thm. 3.1] it has been shown that a non-mergeable product trellis is isomorphic to the corresponding BCJR-trellis. The following theorem generalizes this result. It shows that every (one-to-one) product trellis $T_{G,S}$ can be merged to the corresponding BCJR-trellis and merging is accomplished by taking suitable quotients of the vertex spaces. An instance of this process has been given in Example 4.10. It also reminds us to be aware of the fact that the resulting BCJR-trellis may not be one-to-one.

Theorem 4.14 *Let S be as in (3.3) and consider the trellises $T_{(G,H,S)}$ and $T_{G,S}$ along with the matrices N_i and M_i as in (4.2)/(4.1) and (3.4), respectively. Then*

- (a) *there exist subspaces $W_i \subseteq \text{im } M_i$ such that $\ker M_i \oplus W_i = \ker N_i$,*
- (b) *the quotient trellis T' of $T_{G,S}$ defined as $T' = (V', E')$, where $V' = \cup_{i=0}^{n-1} V'_i$ and $E' = \cup_{i=0}^{n-1} E'_i$ with*

$$V'_i = \text{im } M_i / W_i, \quad E'_i = \{(\alpha M_i + W_i, \alpha G_i, \alpha M_{i+1} + W_{i+1}) \mid \alpha \in \mathbb{F}^k\},$$

is isomorphic to the BCJR-trellis $T_{(G,H,S)}$. In particular, $\mathcal{C}(T') = \mathcal{C}$.

PROOF: (a) First notice that by the very definition of M_i in (3.4) along with Proposition 4.8(b) we have $M_i N_i = N_i$ and thus $\ker M_i \subseteq \ker N_i$. Hence there exists a subspace $W'_i \subseteq \mathbb{F}^k$ such that $\ker M_i \oplus W'_i = \ker N_i$. Define the subspace $W_i := W'_i M_i := \{\alpha M_i \mid \alpha \in W'_i\} \subseteq \text{im } M_i$. Notice that $(M_i)^2 = M_i$, thus M_i defines a projection and hence $\ker M_i \oplus \text{im } M_i = \mathbb{F}^k$. This implies $\ker M_i \cap W_i = \{0\}$. The inclusion $W_i \subseteq \ker N_i$ follows immediately from the identity $M_i N_i = N_i$ along with the fact that $W'_i \subseteq \ker N_i$. Now let $\beta \in \ker N_i$. Then there exist $\alpha \in \ker M_i$ and $\gamma \in W'_i$ such that $\beta = \alpha + \gamma$. But then $\beta = \alpha + \gamma(I - M_i) + \gamma M_i$ and $\gamma M_i \in W_i$ while $\alpha + \gamma(I - M_i) \in \ker M_i$ due to $(I - M_i) M_i = 0$. All this proves (a).

(b) First of all, by virtue of (a) we have $\dim V'_i = \dim(\text{im } M_i) - \dim W_i = k - (\dim \ker M_i + \dim W_i) = k - \dim \ker N_i = \text{rk } N_i$. Moreover, it is easy to see that

$$\psi_i : V'_i \longrightarrow \text{im } N_i, \quad \alpha M_i + W_i \longmapsto \alpha N_i$$

is a well-defined, surjective homomorphism and thus an isomorphism between the vertex spaces V'_i of T' and $\text{im } N_i$ of $T_{(G,H,S)}$. But then it is clear from the very definition of the edge spaces E'_i above and E_i as in Theorem 4.1 that T' and $T_{(G,H,S)}$ are isomorphic as defined in (e) in the introduction. \square

We close this section with the following observation.

Remark 4.15 In this remark we sketch a proof showing that the class of linear³, non-mergeable (not necessarily one-to-one) trellises coincides with the class of BCJR-trellises as in Definition 4.5, but where the rows of G maybe linearly dependent.

First of all, let $G \in \mathbb{F}^{k \times n}$ where $\text{rk } G < k$. It not hard to show that Theorem 3.3 remains valid (with the exception that $T_{G,S}$ is not one-to-one anymore) and that BCJR-trellises as in Theorem 4.1 and 4.5 can be defined in the same way. The resulting trellis $T_{(G,H,S)}$ still represents $\mathcal{C} = \text{im } G = \ker H^T$. Moreover, one can easily verify that Propositions 4.3, 4.8, Corollary 4.9, Theorem 4.11, and Theorem 4.14 remain valid with the same proofs (with the exception of part 4.11(b); see also the footnote on Page 7). For sake of clarity let us call the resulting BCJR-trellises genBCJR-trellises.

Now the above implies in particular that every genBCJR-trellis is non-mergeable. For the

³Recall from Definition 2.1(a) that linear trellises are reduced.

converse, let T be a linear and non-mergeable trellis of \mathcal{C} . By [9, Thm. 4.2] T is the product of elementary trellises $T_{x_l, (a_l, b_l)}$, $l = 1, \dots, m$, where $\{x_1, \dots, x_m\}$ is a (not necessarily linearly independent) generating set of \mathcal{C} and where (a_l, b_l) is a generalized span of x_l , that is, $x_{l,j} = 0$ for $j \notin [a_l, b_l]$ (notice the difference to our definition of a span in Def. 3.1). By [9, Lem. 4.3] each factor $T_{x_l, (a_l, b_l)}$ is non-mergeable. But then it is easy to see that (a_l, b_l) is a span of x_l in the sense of Def. 3.1 (that is, x_l is nonzero at the endpoints a_l and b_l) for otherwise the elementary trellis could be merged at time a_l or b_l . Hence $T = T_{G, \mathcal{S}}$ in the sense of Def. 3.10 but where the rows of G are possibly linearly dependent. By Theorem 4.14 T can be merged to the corresponding genBCJR-trellis, hence, by non-mergeability, T is a genBCJR-trellis. As a consequence, if T is one-to-one, then T is a BCJR-trellis.

5 On a Duality Conjecture by Koetter/Vardy

Koetter/Vardy have shown the following.

Lemma 5.1 ([9, Lem. 5.11]) *Let $\mathcal{C} \subseteq \mathbb{F}^n$ and \mathcal{C}^\perp both be codes with support \mathcal{I} and let the characteristic span list of \mathcal{C} be given by $\mathcal{T} = [(a_l, b_l), l = 1, \dots, n]$. Then the characteristic span list of \mathcal{C}^\perp is given by $[(b_l, a_l), l = 1, \dots, n]$.*

Moreover, in [9, Prop. 5.13] they proved the following: select k linearly independent rows of a characteristic matrix of \mathcal{C} , with spans, say, (a_{l_i}, b_{l_i}) , $i = 1, \dots, k$, and construct the resulting KV-trellis T . Pick the $n - k$ rows of the characteristic matrix of \mathcal{C}^\perp that do not have spans (b_{l_i}, a_{l_i}) , $i = 1, \dots, k$ and construct the resulting product trellis \hat{T} . Then the trellises T and \hat{T} share the same SCP. However, while the trellis T for \mathcal{C} is, by construction, a KV-trellis in the sense of Definition 3.10, the trellis \hat{T} might not even represent \mathcal{C}^\perp because it is not a priori clear whether the underlying $n - k$ rows are linearly independent. Koetter/Vardy conjecture that those rows are indeed always linearly independent and thus the trellis \hat{T} is a KV-trellis for \mathcal{C}^\perp . For all this one has to have in mind that Koetter/Vardy define the characteristic matrix in a way which makes it unique for a given code and their conjecture makes sense as phrased. However, in our definition a characteristic matrix is not unique and therefore the linear independence of the selected rows might even depend on the choice of that matrix. This is indeed the case as the following example shows.

Example 5.2 Consider again Example 3.12. The underlying binary code and its dual are $\mathcal{C} = \{0000, 1100, 0111, 1011\}$ and $\mathcal{C}^\perp = \{0000, 1101, 0011, 1110\}$. A characteristic pair for the dual code \mathcal{C}^\perp is given by

$$(\hat{X}, \hat{\mathcal{T}}) = \left(\begin{pmatrix} 1 & 1 & 0 & 1 \\ 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 \end{pmatrix}, [(1, 0], (3, 1], (0, 2], (2, 3)] \right).$$

In Example 3.12 we observed that the last two rows of X are linearly independent and thus give rise to a KV-trellis of \mathcal{C} . The according spans are $(2, 0]$ and $(3, 2]$. Hence the complementary rows for \mathcal{C}^\perp are the first two rows of \hat{X} , but they are linearly dependent. Replacing the first row of \hat{X} with the codeword $(1, 1, 1, 0) \in \mathcal{C}^\perp$ and keeping the span $(1, 0]$, one obtains another characteristic matrix for \mathcal{C}^\perp for which that row selection is indeed linearly independent. All this shows that the conjecture of Koetter/Vardy does depend on the choice of the characteristic matrix. One might also want to observe that both X and \hat{X} are the lexicographically

first characteristic matrix of their respective code and therefore the ones Koetter/Vardy singled out in their approach. Thus the example also shows that their conjecture does not hold for this particular choice.

For the specific classes of self-dual codes and cyclic codes the Koetter/Vardy conjecture has been proven in [7].

In the sequel we will show that for row selections leading to minimal KV-trellises (in the sense of Definition 2.2(a)) there always exists a characteristic matrix of \mathcal{C}^\perp such that Koetter/Vardy's conjecture holds true.

Theorem 5.3 *Let $\mathcal{C}, \mathcal{C}^\perp \subseteq \mathbb{F}^n$ both have support \mathcal{I} . Furthermore, let (X, \mathcal{T}) be a characteristic pair of \mathcal{C} , where $\mathcal{T} = [(a_l, b_l], l = 1, \dots, n]$. Suppose the rows of X with indices l_1, \dots, l_k are linearly independent and give rise to a minimal KV-trellis. Then there exists a characteristic matrix \hat{X} of \mathcal{C}^\perp such that the $n - k$ rows of \hat{X} that do not have spans $(b_{l_i}, a_{l_i}], i = 1, \dots, k$, are linearly independent and give rise to a minimal KV-trellis of \mathcal{C}^\perp .*

In the proof we will use a dualization technique for trellises going back to Mittelholzer [13] and Forney [5, Sec. VII.D and Thm. 8.4]. Their construction of a dual trellis amounts to dualizing the edge spaces (called local constraints in [5]) with respect to a particular bilinear form. In our case this bilinear form reads as in (5.2) below. As the proof will show for minimal BCJR-trellises the resulting dual trellis coincides with the BCJR-dual defined in Proposition 4.4.

PROOF: Without loss of generality let us assume $(l_1, \dots, l_k) = (1, \dots, k)$ and let $G \in \mathbb{F}^{k \times n}$ be the matrix consisting of the first k rows of X . Then G is a generator matrix of \mathcal{C} . As before, let the parity check matrix H be as in (2.2). By Theorem 4.12 we know that the corresponding KV-trellis $T_{G, \mathcal{S}}$ is isomorphic to the BCJR-trellis $T := T_{(G, H, \mathcal{S})}$, where $\mathcal{S} = [(a_l, b_l], l = 1, \dots, k]$. Denote the sets of starting and end points by $\mathcal{A} = \{a_1, \dots, a_k\}$ and $\mathcal{B} = \{b_1, \dots, b_k\}$. The vertex and edge spaces of T are given by $\text{im } N_i$ and $E_i := \text{im}(N_i, G_i, N_{i+1})$, $i \in \mathcal{I}$, where N_i is as in (4.1), (4.2). Let $s_i := \text{rk}(N_i)$, thus (s_0, \dots, s_{n-1}) is the SCP of T .

By Proposition 4.4 the BCJR-trellis $\hat{T} := T_{(H, G, N_0^\top)}$ represents \mathcal{C}^\perp and has the same SCP. Its vertex and edge spaces are given by $\text{im } \hat{N}_i$ and $\hat{E}_i := \text{im}(\hat{N}_i, H_i^\top, \hat{N}_{i+1})$, where $\hat{N}_i = N_i^\top$. Since T is a minimal trellis the same is true for \hat{T} ; see also [15, Thm. 4.1]. Therefore, by virtue of Theorem 3.14, the trellis \hat{T} must be a KV-trellis for \mathcal{C}^\perp . This means, there exists a characteristic pair $(\hat{X}, \hat{\mathcal{T}})$ of \mathcal{C}^\perp such that \hat{T} is a KV $_{(\hat{X}, \hat{\mathcal{T}})}$ -trellis, which in turn implies that there exists a selection of $n - k$ rows of \hat{X} giving rise to the trellis \hat{T} . Let $\hat{\mathcal{S}}$ be the list of associated characteristic spans and let $\hat{\mathcal{A}}$ be the set of starting points of $\hat{\mathcal{S}}$. Thus, $|\hat{\mathcal{A}}| = n - k$.

By virtue of Lemma 5.1 we have to show that $\hat{\mathcal{A}} = \{b_{k+1}, \dots, b_n\}$. From Theorem 4.11 and Proposition 3.5 we know

$$\hat{e}_i := \dim \hat{E}_i = s_i + I_i^{\hat{\mathcal{A}}} \quad \text{and} \quad e_i := \dim E_i = s_{i+1} + I_i^{\mathcal{B}}, \quad i \in \mathcal{I}. \quad (5.1)$$

Using that $\hat{N}_i = N_i^\top$ one obtains the following well-defined non-degenerate bilinear form

$$\begin{aligned} \Pi : (\text{im } N_i \times \mathbb{F} \times \text{im } N_{i+1}) \times (\text{im } \hat{N}_i \times \mathbb{F} \times \text{im } \hat{N}_{i+1}) &\longrightarrow \mathbb{F} \\ ((\alpha N_i, a, \tilde{\alpha} N_{i+1}), (\beta \hat{N}_i, b, \tilde{\beta} \hat{N}_{i+1})) &\longmapsto \alpha N_i \beta^\top + ab - \tilde{\alpha} N_{i+1} \tilde{\beta}^\top. \end{aligned} \quad (5.2)$$

Then we have for all $\alpha \in \mathbb{F}^k$ and $\beta \in \mathbb{F}^{n-k}$ and $i \in \mathcal{I}$

$$\begin{aligned}\Pi((\alpha N_i, \alpha G_i, \alpha N_{i+1}), (\beta \hat{N}_i, \beta H_i^\top, \beta \hat{N}_{i+1})) &= \alpha N_i \beta^\top + \alpha G_i H_i \beta^\top - \alpha N_{i+1} \beta^\top \\ &= \alpha(N_i + G_i H_i - N_{i+1}) \beta^\top = 0,\end{aligned}$$

where the last identity follows from the very definition of N_{i+1} in (4.1). Hence $\hat{E}_i \subseteq (E_i)^\circ$, where $(E_i)^\circ$ denotes the dual space of E_i in $\text{im } \hat{N}_i \times \mathbb{F} \times \text{im } \hat{N}_{i+1}$ with respect to the bilinear form Π . Using that $\dim(\text{im } \hat{N}_i \times \mathbb{F} \times \text{im } \hat{N}_{i+1}) = \dim(\text{im } N_i \times \mathbb{F} \times \text{im } N_{i+1}) = s_i + s_{i+1} + 1$ and applying the second identity in (5.1), we conclude

$$\hat{e}_i = \dim \hat{E}_i \leq \dim(E_i)^\circ = s_i + s_{i+1} + 1 - e_i = s_i + 1 - I_i^{\mathcal{B}}. \quad (5.3)$$

Along with the first identity in (5.1) this yields

$$I_i^{\hat{\mathcal{A}}} + I_i^{\mathcal{B}} \leq 1 \text{ for all } i \in \mathcal{I} \quad (5.4)$$

for the indicator functions on $\hat{\mathcal{A}}$ and \mathcal{B} . Hence the sets $\hat{\mathcal{A}}$ and \mathcal{B} are disjoint and since the two sets are subsets of \mathcal{I} of cardinality $n - k$ and k , respectively, we may conclude $\hat{\mathcal{A}} = \mathcal{I} \setminus \mathcal{B} = \{b_{k+1}, \dots, b_n\}$ as desired. This completes the proof.

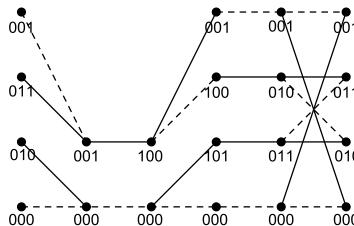
One should notice that we now also have equality in (5.4) and (5.3) and thus derived the identity $\hat{E}_i = (E_i)^\circ$, which is an instance of [5, Cor. 8.2]. \square

It is worth noting that the proof does not explicitly require the minimality of the trellis, but only that the dual trellis $\hat{T} = T_{(H, G, N_0^\top)}$ be a KV-trellis. We strongly believe that Theorem 5.3 is true for all KV-trellises (and not just minimal ones). Moreover, we believe that the BCJR-dual of a KV-trellis is isomorphic to the trellis obtained from the dual row selection of a suitably chosen dual characteristic matrix.

Remark 5.4 The proof above shows that for KV-trellises whose BCJR-dual is a KV-trellis again, this BCJR-dual coincides with the dual obtained by dualizing the edge spaces. In other words, $\hat{E}_i = (E_i)^\circ$ for $i \in \mathcal{I}$ in the notation of the proof above. We wish to point out that this is not always the case for BCJR-trellises, not even if they are one-to-one and isomorphic to the corresponding product trellis (recall that BCJR-trellises are always mergeable). An example can be drawn from Example 4.13. The trellis $T_{G, S} \cong T_{(G, H, S)}$ has SCP $(2, 1, 1, 2, 2)$ and ECP $(2, 2, 2, 3, 2)$. The dual trellis $\hat{T} = T_{(H, G, N_0^\top)}$ has, of course, the same SCP and its ECP is $(2, 1, 2, 2, 2)$. According to (5.3) the trellis T° with vertex spaces $\text{im } \hat{N}_i$ and edge spaces $(E_i)^\circ$ has ECP $(2, 1, 2, 2, 3)$, hence $\hat{E}_4 \subsetneq (E_4)^\circ$ and $\hat{E}_i = (E_i)^\circ$ for $i = 0, \dots, 3$. One computes

$$(E_4)^\circ = \text{im} \begin{pmatrix} 0 & 1 & 1 & | & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & | & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & | & 1 & 0 & 0 & 1 \end{pmatrix} \in \text{im } \hat{N}_4 \times \mathbb{F} \times \text{im } \hat{N}_0,$$

which results in the following trellis



[Trellis T°]

As one can easily see, this trellis does indeed represent \mathcal{C}^\perp , is one-to-one, but it is not reduced! As a consequence, it is not considered a linear trellis due to Definition 2.1(a). Removing the 4 cross-diagonal edges in $(E_4)^\circ$ results in the BCJR-dual $T_{(H,G,N_0^\top)}$ displayed in Example 4.13.

Some Open Problems

- In Theorem 3.14 we have seen that every minimal trellis is a $\text{KV}_{(X,\mathcal{T})}$ -trellis for a particular choice of the characteristic pair (X, \mathcal{T}) . Example 3.15 showed that indeed a suitable choice of the characteristic matrix is needed for this result to be true. In that particular example a different choice of the characteristic matrix leads to a trellis that is (only) structurally isomorphic to the given trellis. All this gives rise to the question whether every minimal trellis arises, up to structural isomorphism, from a fixed characteristic matrix? It is easy to see that this amounts to investigating whether if a certain row selection of one characteristic matrix is linearly independent and gives rise to a minimal trellis of \mathcal{C} , then the same row selection is linearly independent for each characteristic matrix of the code. If this turns out to be true then this implies that in Theorem 5.3 the statement would be true for each characteristic matrix of the dual code.
- Does Theorem 5.3 hold true for all KV-trellises? This would be answered in the affirmative if one can show that the BCJR-dual of a KV-trellis is a KV-trellis again.
- From Examples 3.12/5.2 it is clear that the duality result of Theorem 5.3 does indeed depend on the choice of the characteristic matrix of the dual code (at least, if the trellises are not minimal). However, several examples suggest that one may ask whether there exist a dual pairing of characteristic matrices X and \hat{X} for \mathcal{C} and \mathcal{C}^\perp , respectively, such that Koetter/Vardy’s duality conjecture holds true for these pairs, that is, such that the Koetter/Vardy-dual of a $\text{KV}_{(X,\mathcal{T})}$ -trellis is a $\text{KV}_{(\hat{X},\hat{\mathcal{T}})}$ -trellis. For instance, for Examples 3.12/5.2 this is true for the pairing (X', \hat{X}) and when pairing X with the (unique) remaining characteristic matrix of \mathcal{C}^\perp .
- Compare the BCJR-dualization with the dualization of trellises via dualizing the edge spaces as in [13, 5]. More specifically, under which conditions do these dual trellises coincide when using the bilinear form (5.2)?

References

- [1] L. R. Bahl, J. Cocke, F. Jelinek, and J. Raviv. Optimal decoding of linear codes for minimizing symbol error rate. *IEEE Trans. Inform. Theory*, IT-20:284–287, 1974.
- [2] A. R. Calderbank, G. D. Forney, Jr., and A. Vardy. Minimal tail-biting trellises: The Golay code and more. *IEEE Trans. Inform. Theory*, IT-45:1435–1455, 1999.
- [3] G. D. Forney, Jr. Minimal realizations of linear systems: The “shortest Basis” approach. Preprint 2009. ArXiv: 0910.4336.
- [4] G. D. Forney, Jr. Coset codes II: Binary lattices and related codes. *IEEE Trans. Inform. Theory*, IT-34:1152–1187, 1988.
- [5] G. D. Forney, Jr. Codes on graphs: Normal realizations. *IEEE Trans. Inform. Theory*, IT-47:520–548, 2001.

- [6] G. D. Forney, Jr. and M. D. Trott. The dynamics of group codes: State spaces, trellis diagrams, and canonical encoders. *IEEE Trans. Inform. Theory*, IT-39:1491–1513, 1993.
- [7] H. Kan and H. Shen. A relation between the characteristic generators of a linear code and its dual. *IEEE Trans. Inform. Theory*, IT-51:1199–1202, 2005.
- [8] R. Koetter and A. Vardy. On the theory of linear trellises. In M. Blaum, P. G. Farrel, and H. C. A. van Tilborg, editors, *Information, Coding and Mathematics*, pages 323–354. Kluwer, Boston, 2002.
- [9] R. Koetter and A. Vardy. The structure of tail-biting trellises: Minimality and basic principles. *IEEE Trans. Inform. Theory*, IT-49:2081–2105, 2003.
- [10] F. R. Kschischang and V. Sorokine. On the trellis structure of block codes. *IEEE Trans. Inform. Theory*, IT-41:1924–1937, 1995.
- [11] S. Lin and R. Y. Shao. General structure and construction of tail-biting trellises for linear block codes. In *Proceedings of the 2000 IEEE International Symposium on Information Theory*, page 117, 2000.
- [12] R. J. McEliece. On the BCJR Trellis for linear block codes. *IEEE Trans. Inform. Theory*, IT-42:1072–1092, 1996.
- [13] T. Mittelholzer. Convolutional codes over groups: A pragmatic approach. In *Proc. of the 33rd Allerton Conference on Communications, Control, and Computing*, pages 380–381, 1995.
- [14] D. J. Muder. Minimal trellises for block codes. *IEEE Trans. Inform. Theory*, IT-34:1049–1053, 1988.
- [15] A. V. Nori and P. Shankar. Unifying views of tail-biting trellis constructions for linear block codes. *IEEE Trans. Inform. Theory*, IT-52:4431–4443, 2006.
- [16] J. W. Polderman and J. C. Willems. *Introduction to Mathematical Systems Theory: A behavioral approach*. Springer, New York, 1998.
- [17] Y. Shany and Y. Be'ery. Linear tail-biting trellises, the square root bound, and applications for Reed-Muller codes. *IEEE Trans. Inform. Theory*, IT-46:1514–1523, 2000.
- [18] A. Vardy. Trellis structure of codes. In V. S. Pless and W. C. Huffman, editors, *Handbook of Coding Theory, Vol. 2*, pages 1989–2117. Elsevier, Amsterdam, 1998.
- [19] J. C. Willems. Models for dynamics. *Dynamics Reported*, 2:171–269, 1989.
- [20] J. K. Wolf. Efficient maximum likelihood decoding of linear block codes. *IEEE Trans. Inform. Theory*, IT-24:76–80, 1978.